(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
21 October 2004 (21.10.2004)

PCT

(10) International Publication Number
WO 2004/090146 A1

(51) International Patent Classification[7]: C12N 15/90, 15/79

(21) International Application Number: PCT/FI2004/000228

(22) International Filing Date: 14 April 2004 (14.04.2004)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
20030561 14 April 2003 (14.04.2003) FI

(71) Applicant *(for all designated States except US)*: **FINNZYMES OY** [FI/FI]; Riihitontuntie 14 B, FI-02200 Espoo (FI).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **SAVILAHTI, Harri** [FI/FI]; Munkkiniemen Puistotie 2 B 34, FI-00330 Helsinki (FI). **FRILANDER, Mikko** [FI/FI]; Kekomäentie 20 C, FI-00940 Helsinki (FI). **XIAOJUAN, Meng** [CN/FI]; Kytösuontie 9 as 1, FI-00300 Helsinki (FI). **PAATERO, Anja** [FI/FI]; Taulutie 26 C, FI-00680 Helsinki (FI). **PAJUNEN, Maria** [FI/FI]; Tuomipolku 7 E, FI-00780 Helsinki (FI). **TURAKAINEN, Hilkka** [FI/FI]; Tammihaantie 15-17 E 19, FI-02940 Espoo (FI).

(74) Agent: **OY JALO ANT-WUORINEN AB**; Iso Roobertinkatu 4-6 A, FI-00120 Helsinki (FI).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NA, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— *with international search report*
— *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: DELIVERY OF NUCLEIC ACIDS INTO EUKARYOTIC GENOMES USING IN VITRO ASSEMBLED MU TRANSPOSITION COMPLEXES

(57) Abstract: The present invention relates to genetic engineering and especially to the use of DNA transposition complex of bacteriophage Mu. In particular, the invention provides a gene transfer system for eukaryotic cells, wherein *in vitro* assembled Mu transposition complexes are introduced into a target cell and subsequently transposition into a cellular nucleic acid occurs. The invention further provides a kit for producing insertional mutations into the genomes of eukaryotic cells. The kit can be used, e.g., to generate insertional mutant libraries.

WO 2004/090146 A1

Delivery of nucleic acids into eukaryotic genomes using in vitro assembled mu transposition complexes

The present invention relates to genetic engineering and especially to the use of DNA transposition complex of bacteriophage Mu. In particular, the invention provides a gene transfer system for eukaryotic cells, wherein *in vitro* assembled Mu transposition complexes are introduced into a target cell. Inside the cell, the complexes readily mediate integration of a transposon construct into a cellular nucleic acid. The invention further provides a kit for producing insertional mutations into the genomes of eukaryotic cells. The kit can be used, e.g., to generate insertional mutant libraries.

BACKGROUND OF THE INVENTION

Efficient transfer of nucleic acid into a target cell is prerequisite for the success of almost any molecular biology application. The transfer of nucleic acid into various types of cells provides means to study gene function in living organisms, to express exogenous genes, or to regulate cell functions such as protein expression. Stably transferred inserts can also be used as primer binding sites in sequencing projects. In principle, the transfer can be classified as transient or stable. In the former case the transferred genetic material will eventually disappear from the target cells. Transient gene transfer typically utilizes plasmid constructions that do not replicate within the host cell. Because vector molecules that would replicate in mammalian cells are scarce, and in essence they are limited to those involving viral replicons (i.e. no plasmids available), the transient transfer strategy is in many cases the only straightforward gene transfer strategy for mammalian cells. For other types of cells, e.g. bacterial and lower eukaryotes such as yeast, replicating plasmids are available and therefore transient expression needs to be used only in certain specific situations in which some benefits can be envisioned (e.g. conditional expression).

In many cases stable gene transfer is the preferred option. For bacteria and lower eukaryotes plasmids that replicate within the cells are available. Accordingly, these DNA molecules can be used as gene delivery vehicles. However, the copy numbers of such plasmids typically exceeds one or two and therefore the transferred genes increase the gene dosage substantially. Typically used plasmids for bacteria and yeasts are present in tens or hundreds of copies. Increased gene dosage compared to normal situation is a potential source of artefactual or at least biased experimental results in many systems. Therefore, it

would be advantageous to generate situations in which single-copy gene transfer (per haploid genome) would be possible.

In general, stable single-copy gene transfer can be achieved if transferred DNA can be inserted into the target cell's chromosomal DNA. Traditionally, this has been achieved by using different types of recombination reactions. In bacteria, homologous recombination and site-specific recombination are both widely used and in some cases yet less well characterized "illegitimate" recombination may be used. The choice of a method typically depends on whether a random or targeted mutation is required. While some of these methods are relatively trivial to use for a subset of the bacterial species, a general-purpose method would be more desirable.

Recombination reactions may also be used to stably transfer DNA into eukaryotic cell's chromosomal DNA. Homologous and site-specific recombination reactions produce targeted integrations, and "illegitimate" recombination generates non-targeted events. Utilization of transpositional recombination has been described for baker's yeast *Saccharomyces cerevisiae* (Ji et al 1993) and for fission yeast *Schizosaccharomyces pombe* (Behrens et al 2000). These strategies involve *in vivo* transposition in which the transposon is launched from within the cell itself. They utilize suitably modified transposons in combination with transposase proteins that are produced within a given cell. Similar systems, in which transposase proteins are produced within cells, are available also for other eukaryotic organisms; typical examples include *Drosophila* and Zebra fish (Rubin and Spradling 1982, Raz et al. 1997).

While transposition systems based on *in vivo* expression of the transposition machinery are relatively straightforward to use they are not an optimal choice for gene transfer for various reasons. For example, efficiency as well as the host-range may be limited, and target site selection may not be optimal. Viral systems, especially retroviral insertion methods, have been used to generate genomic insertions for animal cells. These strategies also have some disadvantageous properties. For example, immune response may be elicited as a response to virally-encoded proteins, and in general, constructing safe and efficient virus vectors and respective packaging cell lines for a given application is not necessarily a trivial task. Therefore, also for eukaryotic cells, a general-purpose random non-viral DNA insertion strategy would be desirable. Introduction of in vitro-assembled transposition complexes

into the cells may be a choice. It is likely that utilization of *in vitro*-assembled DNA transposition complexes may be one of the most versatile systems for gene transfer. Recently, such a system for bacterial cells has been described and it utilizes chemical reactions based on transpositional DNA recombination (US 6,159,736 and US 6,294,385). Efficient systems are expected to provide a pool of mutants that can be used various ways to study many types of aspects of cellular life. These mutant pools are essential for studies involving whole genomes (i.e. functional genomics studies). However, *a priori* it is not possible to envision whether *in vitro*-assembled DNA transposition complexes would work when introduced into eukaryotic cells, especially if the components are derived from the prokaryota. The difference between prokaryotic and eukaryotic cells, especially the presence of nuclear membrane and packaging of eukaryotic genomic DNA into chromatin structure, may prevent the prokaryotic systems from functioning. In addition, in view of the stability and catalytic activity of the transposition complex, conditions within eukaryotic cells may be substantially different from prokaryotic cells. In addition, other unknown restriction system(s) may fight against incoming DNA and non-specific proteases may destroy assembled transposition complexes before they execute their function for integration. Furthermore, even if the transpositional reaction integrates the transposon into the genome, the ensuing 5-bp single-stranded regions (and in some cases 4-nt flanking DNA flaps) would need to be corrected by the host. Therefore, it is clear that the stability and efficiency of transposition complexes inside a eukaryotic cell cannot be predicted from the results with bacterial cells as disclosed in US 6,159,736 and US 6,294,385. Thus, to date there is no indication in the prior art that *in vitro*-assembled transposition complexes can generally be used for nucleic acid transfer into the cells of higher organisms (i.e. eukaryotes).

Bacteriophage Mu replicates its genome using DNA transposition machinery and is one of the best characterized mobile genetic elements (Mizuuchi 1992; Chaconas et al., 1996). We utilised for the present invention a bacteriophage Mu-derived *in vitro* transposition system that has been introduced recently (Haapa et al. 1999a). Mu transposition complex, the machinery within which the chemical steps of transposition take place, is initially assembled from four MuA transposase protein molecules that first bind to specific binding sites in the transposon ends. The 50 bp Mu right end DNA segment contains two of these binding sites (they are called R1 and R2 and each of them is 22 bp long, Savilahti et al. 1995). When two transposon ends meet, each bound by two MuA monomers, a

transposition complex is formed through conformational changes. Then Mu transposition proceeds within the context of said transposition complex, i.e., protein-DNA complexes that are also called DNA transposition complexes or transpososomes (Mizuuchi 1992, Savilahti et al. 1995). Functional core of these complexes are assembled from a tetramer of MuA transposase protein and Mu-transposon-derived DNA-end-segments (i.e. transposon end sequences recognised by MuA) containing MuA binding sites. When the core complexes are formed they can react in divalent metal ion-dependent manner with any target DNA and insert the Mu end segments into the target (Savilahti et al 1995). A hallmark of Mu transposition is the generation of a 5-bp target site duplication (Allet, 1979; Kahmann and Kamp, 1979).

In the simplest case, the MuA transposase protein and a short 50 bp Mu right-end (R-end) fragment are the only macromolecular components required for transposition complex assembly and function (Savilahti et al. 1995, Savilahti and Mizuuchi 1996). Analogously, when two R-end sequences are located as inverted terminal repeats in a longer DNA molecule, transposition complexes form by synapsing the transposon ends. Target DNA in the Mu DNA *in vitro* transposition reaction can be linear, open circular, or supercoiled (Haapa et al. 1999a).

To date Mu *in vitro* transposition-based strategies have been utilized efficiently for a variety of molecular biology applications including DNA sequencing (Haapa et al. 1999a; Butterfield et al. 2002), generation of DNA constructions for gene targeting (Vilen et al., 2001), and functional analysis of plasmid and viral (HIV) genomic DNA regions (Haapa et al., 1999b, Laurent et al., 2000). Also, functional genomics studies on whole virus genomes of *potato virus A* and bacteriophage PRD1 have been conducted using the Mu *in vitro* transposition-based approaches (Kekarainen et al., 2002, Vilen et al., 2003). In addition, pentapeptide insertion mutagenesis method has been described (Taira et al., 1999). Recently, an insertional mutagenesis strategy for bacterial genomes has been developed in which the *in vitro* assembled functional transpososomes were delivered into various bacterial cells by electroporation (Lamberg et al., 2002).

E. coli is the natural host of bacteriophage Mu. It was first shown with *E. coli* that *in vitro* preassembled transposition complexes can be electroporated into the bacterial cells whereby they then integrate the transposon construct into the genome (Lamberg *et al.*,

2002). The Mu transpososomes were also able to integrate transposons into the genomes of three other Gram negative bacteria tested, namely, *Salmonella enterica* (previously known as *S. typhimurium*), *Erwinia carotovara*, and *Yersinia enterocolitica* (Lamberg et al. 2002). In each of these four bacterial species the integrated transposons were flanked by a 5-bp target site duplication, a hallmark of Mu transposition, thus confirming that the integrations were generated by DNA transposition chemistry.

SUMMARY OF THE INVENTION

We have developed a gene transfer system for eukaryotic cells that utilizes *in vitro*-assembled phage Mu DNA transposition complexes. Linear DNA molecules containing appropriate selectable markers and other genes of interest are generated that are flanked by DNA sequence elements needed for the binding of MuA transposase protein. Incubation of such DNA molecules with MuA protein results in the formation of DNA transposition complexes, transpososomes. These can be delivered into eukaryotic cells by electroporation or by other related methods. The method described in the present invention expands the applicability of the Mu transposon as a gene delivery vehicle into eukaryotes.

In a first aspect, the invention provides a method for incorporating nucleic acid segments into cellular nucleic acid of a eukaryotic target cell, the method comprising the step of:

delivering into the eukaryotic target cell a Mu transposition complex that comprises (i) MuA transposases and (ii) a transposon segment that comprises a pair of Mu end sequences recognised and bound by MuA transposase and an insert sequence between said Mu end sequences, under conditions that allow integration of the transposon segment into the cellular nucleic acid.

In another aspect, the invention features a method for forming an insertion mutant library from a pool of eukaryotic target cells, the method comprising the steps of:

a) delivering into the eukaryotic target cell a Mu transposition complex that comprises (i) MuA transposases and (ii) a transposon segment that comprises a pair of Mu end sequences recognised and bound by MuA transposase and an insert sequence with a selectable marker between said Mu end sequences, under conditions that allow integration of the transposon segment into the cellular nucleic acid,

b) screening for cells that comprise the selectable marker.

In a third aspect, the invention provides a kit for incorporating nucleic acid segments into cellular nucleic acid of a eukaryotic target cell.

The term "transposon", as used herein, refers to a nucleic acid segment, which is recognised by a transposase or an integrase enzyme and which is essential component of a functional nucleic acid-protein complex capable of transposition (i.e. a transpososome). Minimal nucleic acid-protein complex capable of transposition in the Mu system comprises four MuA transposase protein molecules and a transposon with a pair of Mu end sequences that are able to interact with MuA.

The term "transposase" used herein refers to an enzyme, which is an essential component of a functional nucleic acid-protein complex capable of transposition and which is mediating transposition. The term "transposase" also refers to integrases from retrotransposons or of retroviral origin.

The expression "transposition" used herein refers to a reaction wherein a transposon inserts itself into a target nucleic acid. Essential components in a transposition reaction are a transposon and a transposase or an integrase enzyme or some other components needed to form a functional transposition complex. The gene delivery method and materials of the present invention are established by employing the principles of *in vitro* Mu transposition (Haapa et al. 1999ab and Savilahti et al. 1995).

The term "transposon end sequence" used herein refers to the conserved nucleotide sequences at the distal ends of a transposon. The transposon end sequences are responsible for identifying the transposon for transposition.

The term "transposon binding sequence" used herein refers to the conserved nucleotide sequences within the transposon end sequence whereto a transposase specifically binds when mediating transposition.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1. Mini-Mu transposon integration into the yeast chromosomal or plasmid DNA *in vivo* by *in vitro*-assembled Mu transposition complexes comprising of a tetramer of MuA transposase and a mini-Mu transposon.

Figures 2A and 2B. Schematic representation of the Mu-transposons used in this study with the relevant restriction sites. (2A) Both of the yeast transposons contain *TEF* promoter (P_{TEF}), *kan* marker gene and *TEF* terminator (T_{TEF}) embedded between two 50 bp Mu right end sequences. The kanMX4-p15A-Mu transposon contains the additional p15A replicon. Short arrows denote the binding sites of the primers used for sequencing of the out-cloned flanking sequences. The *Bgl*II sites in the ends are used to excise the transposon from the vector plasmid backbone. (2B) The Mu/LoxP-Kan/Neo transposon for transfecting the mouse ES cells. It contains *kan/neo* marker gene between two Mu right end and LoxP sequences. The *kan/neo* marker includes the prokaryotic and eukaryotic promoters and terminators as explained in Materials and methods.

Figure 3. Mu transposition complex formation with KanMX4-Mu (1.5 kb) and KanMX4-p15A-Mu (2.3 kb) substrates analysed by agarose gel electrophoresis. Substrate DNA was incubated with or without MuA, and the reaction products were analysed in the presence or absence of SDS. Samples were electrophoresed on 2 % agarose gel containing 87 mg/ml of heparin and 87 mg/ml of BSA.

Figures 4A and 4B. Southern blot analysis of the insertions into the yeast genome. Genomic DNA of 17 geneticin-resistant FY1679 clones, resulting from the electroporation of the transposition complexes into yeast cells, was digested with *Bam*HI +*Bgl* II (4A) or *Hin*dIII (4B) and probed with kanMX4 DNA. Lanes 1-17, transposon insertion mutants; C, genomic DNA of original *S. cerevisiae* FY1679 recipient strain as a negative control; P, linearized plasmid DNA containing kanMX4-Mu transposon as a positive control; M, molecular size marker. The sizes of plasmid fragments are shown on the left.

Figures 5A and 5B. Distribution of kanMX4-Mu integration sites on yeast chromosomes (5A) and in the repetitive rDNA region on chromosome 12 (5B). The ovals in (5A) designate the centromer of each chromosome. Integration sites in the diploid strain FY1679 are indicated by bars, and the integration sites in the haploid strain FY-3 by bars

with filled circles. Above the line representing yeast genomic DNA are indicated the transposons that contained the kan gene in the orientation of Watson strand, below the line the transposons are in the Crick strand orientation.

Figure 6. Southern blot analysis of HeLa clones transfected with the transposon complexes. Lanes: **1.** Marker with the following bands: 10 kb, 8 kb, 6 kb, 5 kb, 4 kb, 3 kb, 2.5 kb. **2.** HeLa genomic DNA. **3.** HeLa genomic DNA mixed with purified Mu/LoxP-Kan/Neo transposon (about 2.1 kb). HeLA clones: **4.** RGC13 **5.** RGC14 **6.** RGC15 **7.** RGC16 **8.** RGC23 **9.** RGC24 **10.** RGC26

DETAILED DESCRIPTION OF THE INVENTION

The *in vitro* assembled transposition complex is stable but catalytically inactive in conditions devoid of Mg^{2+} or other divalent cations (Savilahti *et al.*, 1995; Savilahti and Mizuuchi, 1996). After electroporation into bacterial cells, these complexes remain functional and become activated for transposition chemistry upon encountering Mg^{2+} ions within the cells, facilitating transposon integration into host chromosomal DNA (Lamberg *et al.*, 2002). The *in vitro* preassembled transpososomes do not need special host cofactors for the integration step *in vivo* (Lamberg *et al.*, 2002). Importantly, once introduced into cells and integrated into the genome, the inserted DNA will remain stable in cells that do not express MuA (Lamberg *et al.*, 2002).

To study if the Mu transposition system with the *in vitro* assembled transpososomes works also for higher organisms we constructed transposons (antibiotic resistance markers connected to Mu ends), assembled the complexes and tested the transposition strategy and target site selection after electroporation of yeast or mouse cells. The transposons were integrated into the genomes with a 5-bp target site duplication flanking the insertion indicating that a genuine DNA transposition reaction had occurred. These results demonstrate that, surprisingly, the conditions in eukaryotic cells allow the integration of Mu DNA. Remarkably, the nuclear membrane, DNA binding proteins, or DNA modifications or conformations did not prevent the integration. Furhermore, the structure and catalytic activity of the Mu complex retained even after repeated concentration steps. This expands the applicability of the Mu transposition strategy into eukaryotes. The benefit of this system is that there is no need to generate an expression system of the transposition machinery for the organism of interest.

The invention provides a method for incorporating nucleic acid segments into cellular nucleic acid of an isolated eukaryotic target cell or a group of such cells (such as a tissue sample or culture), the method comprising the step of:

delivering into the eukaryotic target cell an *in vitro* assembled Mu transposition complex that comprises (i) MuA transposases and (ii) a transposon segment that comprises a pair of Mu end sequences recognised and bound by MuA transposase and an insert sequence between said Mu end sequences, under conditions that allow integration of the transposon segment into the cellular nucleic acid.

For the method, one can assemble *in vitro* stable but catalytically inactive Mu transposition complexes in conditions devoid of Mg^{2+} as disclosed in Savilahti *et al.*, 1995 and Savilahti and Mizuuchi, 1996. In principal, any standard physiological buffer not containing Mg^{2+} is suitable for the assembly of said inactive Mu transposition complexes. However, a preferred *in vitro* transpososome assembly reaction may contain 150 mM Tris-HCl pH 6.0, 50 % (v/v) glycerol, 0.025 % (w/v) Triton X-100, 150 mM NaCl, 0.1 mM EDTA, 55 nM transposon DNA fragment, and 245 nM MuA. The reaction volume may be for example 20 or 80 microliters. The reaction is incubated at about 30°C for 0.5 - 4 h, preferably 2 h. To obtain a sufficient amount of transposition complexes for delivery into the cells, the reaction is then concentrated and desalted from several assembly reactions. For the yeast transformations the final concentration of transposition complexes compared to the assembly reaction is preferably at least tenfold and for the mouse cell transfections at least 20-fold. The concentration step is preferably carried out by using centrifugal filter units. Alternatively, it may be carried out by centrifugation or precipitation (e.g. using PEG or other types of precipitants).

In the method, the concentrated tranposition complex fraction is delivered into the eukaryotic target cell. The preferred delivery method is electroporation. The electroporation of Mu transposition complexes into bacterial cells is disclosed in Lamberg et al., 2002. However, the method of Lamberg et al cannot be directly employed for introduction of the complexes into eukaryotic cells. As shown below in the Experimental Section, the procedure for electroporation of mouse embryonic stem (ES) cells described by Sands and Hasty (1997) can be used in the method of the invention. A variety of other

DNA introduction methods are known for eukaryotic cells and the one skilled in the art can readily utilize these methods in order to carry out the method of the invention (see e.g. "Electroporation Protocols for Microorganisms", ed. Jac A. Nickoloff, Methods in Molecular Biology, volume 47, Humana Press, Totowa, New Jersey, 1995; "Animal Cell Electroporation and Electrofusion Protocols", ed. Jac A. Nickoloff, Methods in Molecular Biology, volume 48, Humana Press, Totowa, New Jersey, 1995; and "Plant cell Electroporation and Electrofusion Protocols", ed. Jac A. Nickoloff, Methods in Molecular Biology, volume 55, Humana Press, Totowa, New Jersey, 1995). Such DNA delivery methods include direct injections by the aid of needles or syringes, exploitation of liposomes, and utilization of various types of transfection-promoting additives. Physical methods such as particle bombardment may also be feasible.

Transposition into the cellular nucleic acid of the target cell seems to follow directly after the electroporation without additional intervention. However, to promote transposition and remedy the stress caused by the electroporation, the cells can be incubated at about room temperature to 30 °C for 10 min - 48 h or longer in a suitable medium before plating or other subsequent steps. Preferably, a single insertion into the cellular nucleic acid of the target cell is produced.

The eukaryotic target cell of the method may be a human, animal (preferably a mammal), plant, fungi or yeast cell. Preferably, the animal cell is a cell of a vertebrate such as mouse (*Mus musculus*), rat (*Rattus norvegicus*), *Xenopus,* Fugu or zebra fish or an invertebrate such as *Drosophila melanogaster* or *Caenorhabditis elegans*. The plant cell is preferably from *Arabidopsis thaliana*, tobacco or rice. The yeast cell is preferably a cell of *Saccharomyces cerevisiae* or *Schizosaccharomyces pombe*.

The insert sequence between Mu end sequences preferably comprises a selectable marker, gene or promoter trap or enhancer trap constructions, protein expressing or RNA producing sequences. Such constructs renders possible the use of the method in gene tagging, functional genomics or gene therapy.

The term "selectable marker" above refers to a gene that, when carried by a transposon, alters the ability of a cell harboring the transposon to grow or survive in a given growth environment relative to a similar cell lacking the selectable marker. The transposon nucleic

acid of the invention preferably contains a positive selectable marker. A positive selectable marker, such as an antibiotic resistance, encodes a product that enables the host to grow and survive in the presence of an agent, which otherwise would inhibit the growth of the organism or kill it. The insert sequence may also contain a reporter gene, which can be any gene encoding a product whose expression is detectable and/or quantitatable by immunological, chemical, biochemical, biological or mechanical assays. A reporter gene product may, for example, have one of the following attributes: fluorescence (e.g., green fluorescent protein), enzymatic activity (e.g., luciferase, *lacZ*/β-galactosidase), toxicity (e.g., ricin) or an ability to be specifically bound by a second molecule (e.g., biotin). The use of markers and reporter genes in eukaryotic cells is well-known in the art.

Since the target site selection of *in vitro* Mu system is known to be random or nearly random, one preferred embodiment of the invention is a method, wherein the nucleic acid segment is incorporated to a random or almost random position of the cellular nucleic acid of the target cell. However, targeting of the transposition can be advantageous in some cases and thus another preferred embodiment of the invention is a method, wherein the nucleic acid segment is incorporated to a targeted position of the cellular nucleic acid of the target cell. This could be accomplished by adding to the transposition complex, or to the DNA region between Mu ends in the transposon, a targeting signal on a nucleic acid or protein level. Said targeting signal is preferably a nucleic acid, protein or peptide which is known to efficiently bind to or associate with a certain nucleotide sequence, thus facilitating targeting.

One specific embodiment of the invention is the method wherein a modified MuA transposase is used. Such MuA transposase may be modified, e.g., by a deletion, an insertion or a point mutation and it may have different catalytic activities or specifities than an unmodified MuA.

Another embodiment of the invention is a method for forming an insertion mutant library from a pool of eukaryotic target cells, the method comprising the steps of:

a) delivering into the eukaryotic target cell an *in vitro* assembled Mu transposition complex that comprises (i) MuA transposases and (ii) a transposon segment that comprises a pair of Mu end sequences recognised and bound by MuA transposase and an insert sequence with

a selectable marker between said Mu end sequences, under conditions that allow integration of the transposon segment into the cellular nucleic acid.

b) screening for cells that comprise the selectable marker.

In the above method, a person skilled in the art can easily utilise different screening techniques. The screening step can be performed, e.g., by methods involving sequence analysis, nucleic acid hybridisation, primer extension or antibody binding. These methods are well-known in the art (see, for example, Current Protocols in Molecular Biology, eds. Ausubel et al, John Wiley & Sons: 1992). Libraries formed according to the the method of the invention can also be screened for genotypic or phenotypic changes after transposition.

Further embodiment of the invention is a kit for incorporating nucleic acid segments into cellular nucleic acid of a eukaryotic target cell. The kit comprises a concentrated fraction of Mu transposition complexes that comprise a transposon segment with a marker, which is selectable in eukaryotic cells. Preferably, said complexes are provided as a substantially pure preparation apart from other proteins, genetic material, and the like.

The publications and other materials used herein to illuminate the background of the invention, and in particular, to provide additional details with respect to its practice, are incorporated herein by reference. The invention will be described in more detail in the following Experimental Section.

EXPERIMENTAL SECTION

MATERIALS AND METHODS

Strains, cell lines and media

The *Eschericia coli* DH5α was used for bacterial transformations. The bacteria were grown at 37 °C in LB broth or on LB agar plates. For the selection and maintenance of plasmids, antibiotics were used at the following concentrations: ampicillin 100-150 μg/ml, kanamycin 10-25 μg/ml, and chloramphenicol 10 μg/ml. The *Saccharomyces cerevisiae* strain FY1679 (*MATa/MATα ura3-52/ura 3-52 his3Δ200/HIS3 leu2Δ1/LEU2 trp1Δ63/TRP1 GAL2/GAL2*; Winston et al. 1995) and its haploid derivative FY-3 (*MAT*a

HIS LEU TRP ura3-52) were used for yeast transformations. The yeasts were grown on YPD (1 % yeast extract, 2 % peptone, 2 % glucose) or minimal medium (0.67 % yeast nitrogen base, 2 % glucose). For the selection of the transformants, yeast cells were grown on YPD plates containing 200 μg/ml of G418 (geneticin, Sigma).

The procedures required for propagating mouse AB2.2-Prime embryonic stem (ES) cells (Lexicon Genetics, Inc.) have been described by Sands and Hasty (1997). Briefly, undifferentiated AB2.2-Prime ES cells were grown on 0.1 % gelatin (Sigma)-coated tissues culture plates in the ES culture medium consisting of DMEM (Gibco) supplemented with 15 % fetal bovine serum (Hyclone), 2 mM L-glutamine (Gibco), 1 mM Sodium pyruvate (Gibco), 100 μM β-Mercaptoethanol and nonessential amino acids (Gibco), 50 U/ml Penicillin, 50 μg/ml Streptomycin (Gibco), and 1000 U/ml LIF (Chemicon).

HeLa S3 cells (ATCC # CCL-2.2) were grown in cell culture medium consisting of MEM supplemented with 10% fetal bovine serum (Gibco Invitrogen), 2 mM L-glutamine (Gibco Invitrogen), 50 U/ml Penicillin (Gibco Invitrogen), and 50 μg/ml Streptomycin (Gibco Invitrogen).

Proteins and reagents

MuA transposase (MuA), proteinase K, calf intestinal alkaline phosphatase (CIP) and Cam^R Entranceposon (TGS Template Generation System) were obtained from Finnzymes, Espoo, Finland. Restriction endonucleases and the plasmid pUC19 were from New England Biolabs. Klenow enzyme was from Promega. Enzymes were used as recommended by the suppliers. Bovine serum albumin was from Sigma. [α^{32}P]dCTP (1000-3000 Ci/mmol) was from Amersham Biosciences.

Construction of kanMX4-Mu transposons

The *kanMX4* selector module (1.4 kb) was released from the pFA6-kanMX4 (Wach et al. 1994) by *Eco*RI + *Bgl*II double digestion and ligated to the 0.75 kb vector containing the pUC miniorigin and the Mu ends, producing the kanMX4-Mu plasmid, pHTH1. Plasmid DNA was isolated with the Plasmid Maxi Kit (QIAGEN). To confirm the absence of mutations in the *kanMX4* module the insert was sequenced following the *in vitro*

transposition reaction with the Cam^R Entranceposon as a donor DNA and the plasmid pHTH1 as a target DNA with primers Muc1 and Muc2.

The primers for sequencing the yeast constructs were Muc1:
5'-GCTCTCCCCGTGGAGGTAAT-3' (SEQ ID NO:1) and Muc2:
5'-TTCCGTCACAGGTATTTATTCGGT-3' (SEQ ID NO:2).

We also constructed a transposon with a bacterial replicon between the Mu ends to allow easier outcloning. The p15A replicon was cut from the plasmid pACYC184 (Rose 1988) with *Sph*I, blunted with Klenow enzyme, and ligated into *Eco*RI-cut end-filled pHTH1 to produce kanMX4-p15A-Mu plasmid, pHTH4.

Construction of Mu/LoxP-Kan/Neo transposon

A neomycin-resistance cassette containing a bacterial promoter, SV40 origin of replication, SV40 early promoter, kanamycin/neomycin resistance gene, and Herpes simplex virus thymidine kinase polyadenylation signals was generated by PCR from pIRES2-EGFP plasmid (Clontech). After addition of LoxP sites and Mu end sequences using standard PCR-based techniques, the construct was cloned as a *Bgl*II fragment into a vector backbone derived from pUC19. The construct (pALH28) was confirmed by DNA sequencing.

Assembly and concentration of transpososomes

The transposons (kanMX4-Mu, 1.5 kb; kanMX4-p15A-Mu, 2.3 kb; Mu/LoxP-Kan/Neo, 2.1 kb) were isolated by *Bgl*II digestion from their respective carrier plasmids (pHTH1, pHTH4, pALH28). The DNA fragments were purified chromatographically as described (Haapa et al. 1999a).

The standard *in vitro* transpososome assembly reaction (20 µl or 80 µl) contained 55 nM transposon DNA fragment, 245 nM MuA, 150 mM Tris-HCl pH 6.0, 50 % (v/v) glycerol, 0.025 % (w/v) Triton X-100, 150 mM NaCl, 0.1 mM EDTA. The reaction was carried out at 30°C for 2 h. The complexes were concentrated and desalted from several reactions by Centricon concentrator (Amicon) according to manufacturer's instructions and washed once with water. The final concentration for the yeast transformations was approximately tenfold and for the mouse transfections about 20-fold.

Electrocompetent bacterial and yeast cells

Electrocompetent bacterial cells for standard cloning were prepared and used as described (Lamberg et al., 2002). Electrocompetent *S. cerevisiae* cells were grown as follows. An overnight stationary phase culture was diluted 1:10 000 in fresh YPD (1 % yeast extract, 2 % peptone, 2 % glucose) and grown to A_{600} 0.7 – 1.2. The cell pellets were collected by centrifugation (5000 rpm), suspended in ¼ volume of 0.1 M lithium acetate, 10 mM dithiotreitol, 10 mM Tris-HCl pH 7.5, 1 mM EDTA (LiAc/DTT/TE) and incubated at room temperature for 1 h. The repelleted cells were washed with ice-cold water and again collected by centrifugation. The pellet was then resuspended in 1/10 of the original volume of ice-cold 1 M sorbitol. Following centrifugation, the pellet was suspended in ice-cold 1 M sorbitol to yield ~200-fold concentration of the original culture density. One hundred microliters of cell suspension were used for each electroporation. For competence status determionation, transpososomes or plasmid DNA were added to the cell suspension and incubated on ice for 5 min. The mixture was transferred to a 0.2 cm cuvette and pulsed at 1.5 kV (diploid FY 1679) or 2.0 kV (haploid FY-3), 25 μF, 200 ohms with Bio-Rad Genepulser II. After electroporation 1 ml of YPD was added, and cultures were incubated at 30°C for 0-4 hours. Subsequently cells were plated on YPD plates containing 200 μg/ml of G418. The competent status of the yeast strains was evaluated in parallel by electroporation of a control plasmid pYC2/CT (URA3, CEN6/ARSH4, amp^R, pUC ori, Invitrogen) and plating the cells on minimal plates.

Mouse ES cell transfection and colony isolation

The procedures used for electroporation of mouse AB2.2-Prime embryonic stem (ES) cells have been described by Sands and Hasty (1997). Briefly, the AB2.2-Prime ES cells were collected in phosphate-buffered saline (PBS) at a density of $11x10^6$ cells/ml. 2.2-2.3 μg of the transposon complexes or linearized DNA was added to an 0.4 cm electroporation cuvette. For each electroporation, 0.9 ml of ES cell suspension (approximately 10×10^6 cells) was mixed with transpososomes or linear DNA. Electroporation was carried out using Bio-Rad's Gene Pulser and Capacitance Extender at 250 V, 500 μF. After electroporation the cells stood at RT for 10 min and were then plated in gelatin coated plates.The electroporated ES cells were cultured in the conditions mentioned above for 24-48 hours before adding G418 (Gibco) to a final concentration of 150 μg/ml to select transposon insertions. Transfected colonies of ES cells were picked after 10 days in

selection and individual colonies were cultured in separate wells of the 96-wells or 24-wells plates using the conditions described above.

HeLa cell transfection and colony isolation

The HeLa cells were electroporated basically according to the instructions by ATCC. Briefly, the HeLa cells were collected in phosphate-buffered saline (PBS) at a density of 1.8×10^6 cells/ml. 2 – 2.3 µg of the transposon complexes or linearized transposon DNA was added to an 0.4 cm electroporation cuvette. For each electroporation, 0.9 ml of HeLa cell suspension (approximately 1.6×10^6 cells) was mixed with transpososomes or linear DNA. Electroporation was carried out using Bio-Rad's Gene Pulser and Capacitance extender at 250 V, 500µF. After electroporation the cells stood at RT for 10 min and were then plated. The electroporated cells were then cultured in the conditions mentioned above for 60 hours before adding G418 (Gibco Invitrogen) to a final concentration of 400 µg/ml to select transposon insertions. Transfected colonies of HeLa cells were picked after 10-11 days in selection and individual colonies were cultured first in separate wells of the 96-wells plate, and transferred later to separate wells of 24-wells or 6-wells plates and 10 cm plates using the conditions described above.

Isolation of genomic DNA

Yeast Genomic DNA of each geneticin resistant yeast clone was isolated either with QIAGEN Genomic DNA Isolation kit or according to Sherman et al., 1981.

Mouse ES cells Genomic DNA was isolated from ES cell essentially according to the method developed by Miller et al. (1988). ES cells were collected from individual wells from the 24-well cultures and suspended to 500 µl of the proteinase K digestion buffer (10 mM Tris-HCl (pH 8.0), 400 mM NaCl, 10 mM EDTA, 0.5 % SDS, and 200 µg/ml proteinase K). The proteinase K treatment was carried out for 8-16 hours at 55°C. Following the proteinase K treatment 150 µl of 6 M NaCl was added followed by centrifugation at microcentrifuge (30 min, 13 K). The supernatant was collected and precipitated with ethanol to yield DNA pellet that was washed with 70% ethanol and air-dried. DNA was dissolved in TE (10 mM Tris-HCl, pH 8.0 and 1 mM EDTA) buffer.

HeLa cells Genomic DNA was isolated from HeLa cells essentially according to the method developed by Miller et al. (1988). HeLa cells were collected from three 10 cm

plates and suspended to 15 ml of proteinase K digestion buffer (10 mM Tris-HCl (pH 8.0), 400 mM NaCl, 10 mM EDTA, 0.5% SDS, and 200-400 µg/ml proteinase K). The proteinase K treatment was carried out at 55°C for 16-48 hours or until no cells were visible. RNase was added at 25-50 µg/ml and incubated at 37°C for 8-24 hours. Following the RNase treatment 4.5 ml of 6 M NaCl was added followed by centrifugation (SS-34, 11.6-14 K, 20-30 min, 4°C). The supernatant was collected and precipitated with ethanol to yield DNA pellet that was washed with 70% ethanol and air-dried. DNA was dissolved in TE (10 mM Tris-HCl (pH 8.0) and 1 mM EDTA) buffer.

Southern blot

Yeast The DNA was digested with appropriate enzymes. The fragments were electrophoresed on a 0.8 % agarose gel and blotted onto Hybond N+ membrane (Amersham). Southern hybridisation was carried out with [α^{32}P]dCTP -labelled (Random Primed, Roche) kanMX4 (*Bgl*II-*Eco*RI fragment) as a probe.

Mouse ES cells DNA Southern blot hybridisation was performed according to standard methods as described (Sambrook, et al., 1989). 10-15 µg of the wild type and transfected AB2.2-Prime ES cell DNAs were digested with various restriction enzymes and separated on 0.8% agarose gels. The DNA was transferred to a nylon filter (Hybond N+, Amersham) and fixed with UV (Stratalinker, Statagene). Inserted DNA was visualized by hybridisation with a [α-^{32}p] dCTP-labeled (RediprimeII, Amersham) DNA probes (Mu/LoxP-Kan/Neo *BamH*I fragment). Hybridisation was performed at 65°C for 16 hours in solutions containing 1.5 x SSPE, 10% PEG 6000, 7% SDS, 100 µg/ml denatured herring sperm DNA. After the hybridisation, the filter was washed twice 5 min and once 15 min in 2xSSC, 0.5% SDS at 65°C and once or twice for 10 – 15 min in the 0.1xSSC, 0.1%SDS at 65°C. The filter was exposed to a Fuji phosphoimager screen for 8-16 hours and processed in a FujiBAS phosphoimager.

HeLa cells Southern blot hybridisation was performed according to standard methods as described (Sambrook et al., 1989). 10 µg of the wild type and transfected HeLa cell DNAs were digested with *NheI* + *SpeI* and separated on 0.8% agarose gel. The DNA was transferred to a nylon filter (Hybond N+, Amersham) and fixed with UV (Stratalinker, Stratagene). Inserted transposon DNA was visualized by hybridisation with a [α-^{32}P] dCTP-labeled (RediprimeII, Amersham) DNA probe (Mu/LoxP-Kan/Neo transposon).

Hybridisation was performed at 65°C for 16 hours in solutions containing 1.5 x SSPE, 10% PEG 6000, 7% SDS, 100 µg/ml denatured herring sperm DNA. After the hybridisation, the filter was washed three times for 20-40 min in 2 x SSC, 0.5% SDS at 65°C and three times for 20-40 min in 0.1 x SSc, 0.1% SDS at 65°C. The filter was exposed to a Fuji phosphoimager screen for 8-16 hours and processed in a FujiBAS phosphoimager

Determination of target site duplication

Cloning. Yeast genomic DNA was digested with *Bam*HI + *Bgl*II, *Sal*I+ *Xho*I or *Pvu*II to produce a fragment with a transposon attached to its chromosomal DNA flanks. These fragments were then cloned into pUC19 cleaved with *Bam*HI, *Sal*I or *Sma*I, respectively, selecting for kanamycin and ampicillin resistance. Alternatively, clones transfected with kanMX4-p15A were cleaved with *Bam*HI + *Bgl*II, ligated, electroporated and selected for resistance produced by the transposon containing fragments. DNA sequences of transposon borders were determined from these plasmids using transposon specific primers SeqA and SeqMX. Genomic locations were identified using the BLAST search at SGD (Saccharomyces Genome Database; http://genome-www.stanford.edu/Saccharomyces/) or SDSC Biology WorkBench (http://workbench.sdsc.edu/) servers.

The primers for sequencing the ends of cloned yeast inserts were Seq A: 5'-ATCAGCGGCCGCGATCC-3' (SEQ ID NO:3) and Seq MX4: 5'-GGACGAGGCAAGCTAAACAG-3' (SEQ ID NO:4).

PCR amplification. Two micrograms of yeast genomic DNA was digested with *Bam*HI +*Bgl*II or *Nhe*I + *Spe*I. Specific partially double-stranded adapters were made by annealing 2 µM adapter primer 1 (WAP-1) with complementary 2 µM adapter primer 2 (WAP-2*), 3 (WAP-3*), or 4 (WAP-4*). The 3' OH group of the WAP-2*, WAP-3*, and WAP-4* primers was blocked by a primary amine group and the 5' ends were phosphorylated. The restriction fragments (200 ng) generated by *Bam*HI + *Bgl*II were ligated with 22 ng of adapter that was made by annealing primers WAP-1 and WAP-2*, whereas the restriction fragments generated with *Nhe*I + *Spe*I were ligated with the 22 ng of adapter made by annealing primers WAP-1 and WAP-3*. One fifth of the ligation reaction was used as a template to perform PCR amplification at 20 µl to enrich for DNA fragments between the adapter and the transposon with primers Walker-1 and TEFterm-1 or Walker-1 and TEFprom-1. PCR conditions were 94°C, 1 min, 55 °C, 1 min, 72 °C, 4 min for 30 cycles.

Nested PCR was carried out at 50 µl using 2 µl of one hundred-fold diluted primary PCR products as a template using primers Walker-2 and TEFterm-2 or Walker-2 and TEFprom-2 for PCR products produced from *Bam*HI + *Bgl*II fragments and Walker-3 and TEFterm-2 or Walker-3 and TEFprom-2 for PCR products produced from the *Nhe*I + *Spe*I fragments. The PCR conditions were as before. The amplified nested PCR products were sequenced using sequencing primer Mu-2.

One microgram of mouse genomic DNA was digested with *Bgl*II + *Bcl*I or *Nhe*I + *Spe*I. Specific partially double-stranded adapters were made as for the yeast. The restriction fragments (400 ng) generated by *Bcl*I + *Bgl*II were ligated with 44 ng of adapter that was made by annealing primers WAP-1 and WAP-2*, whereas the restriction fragments (200 ng) generated with *Nhe*I + *Spe*I were ligated with the 22 ng of adapter made by annealing primers WAP-1 and WAP-3*. Respectively, one fourth or one fifth of the ligation reaction was used as a template to perform PCR amplification at 20 µl to enrich for DNA fragments between the adapter and the transposon with primers Walker-1 and HSP430 or Walker-1 and HSP431. PCR conditions were 94°C, 1 min, 55 °C, 1 min, 72 °C, 4 min for 30 cycles. Nested PCR was carried out at 50 µl using 2 µl of eighty fold or one hundred-fold diluted primary PCR products as a template using primers Walker-2 and HSP429 or Walker-2 and HSP432 for PCR products produced from *Bcl*I + *Bgl*II fragments and Walker-3 and HSP429 or Walker-3 and HSP432 for PCR products produced from the *Nhe*I + *Spe*I fragments. The PCR conditions were as before. The amplified nested PCR products were sequenced using sequencing primer Mu-2.

Primers for PCR-based detection:

WAP-1 CTAATACCACTCACATAGGGCGGCCGCCCGGGC (SEQ ID NO:5)

WAP-2* GATCGCCCGGGCG-NH2 (SEQ ID NO:6)

WAP-3* CTAGGCCCGGGCG-NH2 (SEQ ID NO:7)

WAP-4* AATTGCCCGGGCG-NH2 (SEQ ID NO:8)

Walker-1 CTAATACCACTCACATAGGG (SEQ ID NO:9)

Walker-2	GGGCGGCCGCCCGGGCGATC (SEQ ID NO:10)
Walker-3	GGGCGGCCGCCCGGGCCTAG (SEQ ID NO:11)
Walker-4	GGGCGGCCGCCCGGGCAATT (SEQ ID NO:12)
TEFterm-1	CTGTCGATTCGATACTAACG (SEQ ID NO:13)
TEFterm-2	CTCTAGATGATCAGCGGCCGCGATCCG (SEQ ID NO:14)
TEFprom-1	TGTCAAGGAGGGTATTCTGG (SEQ ID NO:15)
TEFprom-2	GGTGACCCGGCGGGGACGAGGC (SEQ ID NO:16)
Mu-2	GATCCGTTTTCGCATTTATCGTG (SEQ ID NO:17)
HSP429	GGCCGCATCGATAAGCTTGGGCTGCAGG (SEQ ID NO:18)
HSP430	ACATTGGGTGGAAACATTCC (SEQ ID NO:19)
HSP431	CCAAGTTCGGGTGAAGGC (SEQ ID NO:20)
HSP432	CCCCGGGCGAGTCTAGGGCCGC (SEQ ID NO:21)

HeLa cells The genomic HeLa cell DNA was digested with *BamHI* + *BclI* to produce a fragment with a transposon attached to its chromosomal DNA flanks. These fragments were then cloned into pUC19 cleaved with BamHI, selecting for kanamycin and ampicillin resistance. DNA sequences of transposon borders were determined from these plasmids using transposon specific primers HSP430 and HSP431. Genomic locations were identified using the SSAHA search at Ensembl Human Genome Browser Release 20.34c.1 which is based on the NCBI 34 assembly of the human genome.

RESULTS

Transposon construction and its introduction to the cells

To study if the Mu transposition system works also for eukaryotes (Figure 1) we constructed a kanMX4-Mu transposon containing the kan^R gene from Tn*903* and translational control sequences of the TEF gene of *Ashbya gossypii* between the Mu ends, with or without additional bacterial p15A replicon between the Mu ends (Figure 2A). We studied the assembly of Mu transpososomes by incubating MuA protein with the kanMX4-

Mu transposon and detected stable protein-DNA complexes by agarose gel electrophoresis (Figure 3). The reactions with kanMX4-Mu and kanMX-p15A-Mu transposons produced several bands of protein-DNA complexes that disappeared when the sample was loaded in the presence of SDS indicating that only non-covalent protein-DNA interactions were involved in the complexes. An aliquot of assembly reactions with and without MuA transposase were electroporated into *Saccharomyces cerevisiae* cells and the yeasts were scored for geneticin resistance. The competent status of the yeast strains was evaluated in parallel by electroporation of a control plasmid pYC2/CT. The electroporation efficiency with the transpososomes into the yeast was approximately three orders of magnitude lower than the efficiency with the plasmid (Table 1). This result is consistent with previous results with bacteria (Lamberg et al 2002). Only the sample containing detectable protein-DNA complexes yielded geneticin resistant colonies.

For mouse experiments we constructed a Mu/loxP-Kan/Neo transposon that contained bacterial and eukaryotic promoters, kanamycin/neomycin resistance gene, and Herpes simplex virus thymidine kinase polyadenylation signals (Figure 2B). The transfection of the mouse ES cells with the transpososome resulted in 1720 G418 resistant colonies per μg DNA and the linear control in 330 resistant colonies per μg DNA. Thus the transfection with the transpososome yielded over 5 times more resistant colonies per μg DNA. The control cells with no added DNA did not produce any resistant colonies.

In HeLa cells, transfection with the transpososomes resulted in about 10^3 resistant colonies per μg DNA and transfection with the linear control DNA resulted in about 10^1 resistant colonies per μg DNA. Thus the transpososomes were significantly more efficient in generating transfectants. The control cells with no added transposon did not produce any resistant colonies.

Integration of the transposon into the genome

Southern blot analysis can be used to study whether the transposon DNA was inserted into the genomic DNA of the geneticin-resistant colonies. Digestion of genomic DNA with enzyme(s) which do not cut the transposon produces one fragment hybridising to the transposon probe, and digestion with an enzyme which cuts the transposon once produces two fragments in the case of genuine Mu transposition. Genomic DNA from 17 kanMX4-

Mu transposon integration yeast clones was isolated, digested with *Bam*HI + *Bgl*II that do not cut the transposon sequence, or with *Hin*dIII that cleaves the transposon sequence once and analysed by Southern hybridisation with kanMX4 fragment as the probe. Fifteen isolates generated a single band with a discrete but different gel mobility after *Bam*HI + *Bgl*II digestion (Figure 4A) and two bands after *Hin*dIII digestion (Figure 4B). Control DNA from the recipient strain FY1679 did not generate detectable bands in the analyses. Two isolates (G5 and G14) gave several hybridising fragments after *Bam*HI + *Bgl*II digestion suggesting possibility of multiple transposon integrations. However, these two isolates gave three fragments after *Hin*dIII digestion, instead of doubling the amount of fragments detected in the *Bam*HI + *Bgl*II digestion expected in case of multiple transposon integrations. The sizes of the *Hin*dIII fragments of the isolates G5 and G14 (4.3, 2.4 and 1.3 kb) and the pattern of bands in *Bam*HI + *Bgl*II digestion suggested that the transposon was integrated into the yeast 2μ plasmid (for confirmation of this see sequencing results below). Genomic DNA from 17 G418-resistant isolates of the haploid strain FY-3 was analysed in a similar way after *Xho*I + *Sal*I digestion (which do not cut the transposon) and *Pst*I digestion (one cut in the transposon). Thirteen isolates gave one band after *Xho*I + *Sal*I digestion and two bands after *Pst*I digestion suggesting a single integration. Four isolates gave similar pattern of bands as isolates G5 and G14 of strain FY1679 suggesting integration into the 2μ plasmid (results not shown). These data indicate that in most of the studied clones the transposon DNA was integrated as a single copy into the yeast chromosome. In the rest of the clones a single integration was detected in an episome.

Seven mouse ES cell clones were analysed by Southern blotting. Their chromosomal DNA was digested with *BamH*I which releases almost an entire transposon from the genome. All the clones studied had a band at the same position as the *BamH*I digested pALH28 used as a control. The intensity of the band was similar for all clones studied and for control DNA representing same molar amount of DNA as the genomic samples. This suggests that only one copy of the transposon was integrated into each genome.

In HeLa cells, Southern blot analysis was used to confirm that the G418 resistant colonies had the transposon integrated into their genomes. Digestion of the genomic DNA with restriction enzyme(s) that do not cut the transposon produces one fragment hybridising to the transposon probe. Seven HeLa cell transfectant clones were analysed by Southern blot as shown in Figure 6. Their chromosomal DNA was digested with *NheI* + *SpeI* which do

not cut the transposon. A single band was detected from each of the clones indicating that a single copy of the transposon DNA has been integrated in each of the genomes.

The location of insertions in the chromosomes

Yeast Mu transposons integrate almost randomly into the target DNA (Haapa-Paananen et al., 2002). To test the location and distribution of the transposon insertions we cloned transposon-genomic DNA borders from more than one hundred yeast transformants and sequenced the insertion sites on both sides of the transposon using transposon-specific primers (Seq A + Seq MX4). Exact mapping of the insertion sites was possible by BLAST comparison with the SGD database. We used the strain FY1679 which was used in the yeast whole genome sequencing (Winston et al. 1995) to ensure the correct mapping. The overall distribution of 140 integrations on the 16 chromosomes of the yeast is shown in Figure 5A. All chromosomes were hit at least once. Both ORFs and intergenic regions had transposon integrations (Table 2). List of integrations into the genome is presented in Table 3. In the haploid genome, integrations on the essential genes were naturally missed due to the inviability of the cells. On chromosome XII there seems to be a real "hotspot" for transposon integration but this is an artefact since the "hotspot" is in the approximately 9 kb region encoding ribosomal RNA (Figure 5B). This loci is repeated tandemly 100-200 times in the chromosome XII. In this region, the integrations are distributed randomly. The chromosomes in Figure 5A are drawn according to SGD which shows only two copies of this repeated region (when the systematic sequencing of the yeast genome was done, only two rDNA repeats were sequenced) instead of 100 to 200 copies actually present in a yeast genome consisting of 1 to 2 Mb of DNA. Only nine integrations were found at a distance less than 1 kb from a tRNA gene which shows that Mu-transposon integration differs from that of Ty1-Ty4 elements. Integration closest to the end of a chromosome was 6.3 kb showing the difference to the telomere-preferring Ty5 element. The mean interval distance of insertions was 135 kb and was nowhere near covering the whole genome as a library. However, the distribution was even enough to show the randomness of the integration.

Mouse The sequenced transposon-genomic DNA borders were compared to the Mouse Genome Assembly v 3 using Ensembl Mouse Genome Server. The clone RGC57 contained an integrated transposon in the chromosome 3, duplicating positions 59433906-10, which are located in the last intron of both the ENSMUSESTG00000010433 and

10426. Sequencing showed presence of this 5-bp sequence (target site duplication) on both sides of the integrated transposon.

HeLa cells We cloned transposon-genomic DNA borders from three transfectants and sequenced the insertion sites on both sides of the transposon using transposon-specific primers (HSP 430 and HSP431). The integrations are presented in Table 5. All of these 3 transfectants had intact transposon ends with the 5 bp duplication of the target site at both sides of the transposon.

Integration of the transposon in the yeast 2μ plasmid

Most *S. cerevisiae* strains contain an endogenous 2μ plasmid. The yeast 2μ plasmid is a 6318 bp circular species present extrachromosomally in *S. cerevisiae* at 60-100 copies per cell. The plasmid molecules are resident in the nucleus as minichromosomes with standard nucleosome phasing (Livingston and Hahne 1979; Nelson and Fangman 1979; Taketo et al., 1980).

In 23 clones out of 131 clones (17.6 %) the transposon had integrated in the 2μ plasmid and in 108 clones (82.4 %) the transposon had integrated into the chromosomes in the diploid strain FY1679. In the haploid strain FY-3, four clones out of 49 clones (8.2 %) had the transposon in the 2μ plasmid and 45 clones (91.8 %) had the transposon in the chromosomes.

Transposon target site

Genuine Mu transposition produces a 5-bp target site duplication flanking the integrated transposon (Haapa et al. 1999b). The transposon was flanked by target site duplication in 121 clones (out of 122; 99.2 %) of the strain FY1679 and in 42 clones (out of 46; 91.3 %) in the haploid strain FY-3 confirming that a majority of integrations were generated by DNA transposition chemistry. A consensus sequence of 5 bp duplication (5'-N-Y-G/C-R-N-3') has been observed in both *in vivo* and *in vitro* transposition reactions, the most preferred pentamers being 5'-C-Y-G/C-R-G- 3' (Mizuuchi and Mizuuchi 1993; Haapa-Paananen et at. 2002; Butterfield et al. 2002). In this study, the distribution of nucleotides in duplicated pentamers agreed well with the previous results (Table 4).

Table 1. Number of geneticin-resistant colonies detected following electroporation of transpososomes into yeast strains, cfu/ μg DNA

DNA	FY1679	FY-3
KanMX-Mu + MuA	351	178
KanMX-Mu - MuA	0	1
KanMX-p15A-Mu + MuA	43	61
KanMX-p15A-Mu - MuA	0	0
Plasmid pYC2/CT[a]	6.9×10^5	5.6×10^5

[a] Electroporation of plasmid pYC2/CT DNA served as a control for competent status.

Table 2. Distribution of transposon integrations in FY1679 (diploid) and FY-3 (haploid) strains.

Integration site	FY1679	FY-3	Total
Chromosomal DNA			
Protein coding region			53
Essential gene	12 (1 intron)	0	
Nonessential gene	29	11	
rRNA	12	7	19
tRNA (intron)	1	0	1
Ty	2	0	2
Solo-LTR	1	2	3
Intergenic region	48	23	71
2μ plasmid			
Protein coding region	4	2	6
Intergenic region	12	2	14
	121	47	169

Table 3A. Transposon integration sites and target site duplications in *Saccharomyces cerevisiae* diploid strain FY1679.

	←seqmx4	seqA→	Location*
G1	caacatctagCTCAG	(KanMX4-Mu) CTCAGtgagttccga	chr13:908424-908428
G2	agtactaccaTTGAA	(KanMX4-Mu) TTGAAtttacgttca	chr9:279340-279344
G3	taaaaattcaGGCAT	(KanMX4-Mu) GGCATatacaattat	chr16:569334-568338
G4	taaaccaccaTCTGT	(KanMX4-Mu) TCTGTcgcccatctt	chr12:239388-239392
G5	ctgattactaGCGAA	(KanMX4-Mu) GCGAAgctgcgggtg	2μ:3447-3451(NC_001398)
G6	aagaaaagctCAGTG	(KanMX4-Mu) CAGTGgaataatttt	chr4: 825525-825529
G7	gaactctttcCCCAC	(KanMX4-Mu) CCCACcgatccattg	chr16: 862127-862131
G8	aaagatgaaaCCGAG	(KanMX4-Mu) CCGAGtaagctgcta	chr3: 263950-263954
G9	caatgcatcaTCTAC	(KanMX4-Mu) TCTACattacaaacc	chr2: 766314-766318
G10	tttgttcacgCGGGC	(KanMX4-Mu) CGGGCcgcagttgtg	chr11: 308515-308519
G11	atctgtattaACTTC	(KanMX4-Mu) ACTTCgaggtagtaa	chr7: 854983-854987
G12	ttttcatgttCCTAT	(KanMX4-Mu) CCTATtcttgttctt	chr5: 327111-327115
G13	tatccacttcTTAGA	(KanMX4-Mu) TTAGAgggactatcg	chr12: 456350-456354
G14	aaactgttttACAGA	(KanMX4-Mu) ACAGAtttacgatcg	2μ: 2720-2724
G15	tggagttaggCTGGC	(KanMX4-Mu) CTGGCtcggactggc	chr10: 702930-702934
G16	gagcttctgcTTCAC	(KanMX4-Mu) TTCACgtttttgga	chr7: 568606-568610
G17	taacgctagaGGGGC	(KanMX4-Mu) GGGGCaagaaggaag	chr1: 136875-136879
G18	tccaaccgtaGTGGT	(KanMX4-Mu) GTGGTtatataataa	chr 10: 241383-241387
G19	gggggcaatgGTGAA	(KanMX4-Mu) GTGAAatttcgacgc	chr4:276367-276371
G20	taagagcttgTCCGC	(KanMX4-Mu) TCCGCttcgccccaa	chr13: 904363-904367
G21	cataagtgtaAGCCA	(KanMX4-Mu) AGCCAtatgttccct	chr9: 249583-249587
G22	tctggcttaaACCAG	(KanMX4-Mu) ACCAGcactatgtat	chr4: 544898-544902
G23	gttgaatcttCCGAT	(KanMX4-Mu) CCGATaccatcgaca	chr12: 65144-65148
G34	ccctagcgccTAGGG	(KanMX4-Mu) TAGGGtcgagtactg	chr9:138283-138287
G36	ttgctttaacTAGGA	(KanMX4-Mu) TAGGAaagaataaga	chr15:892270-892274
G37	agagactgaaGACGA	(KanMX4-Mu) GACGAggaaatcaaa	chr16:67656-69660
G38	atggatggcgCTCAA	(KanMX4-Mu) CTCAAgcgtgttacc	chr12:453865-453869
G40	tccatcttctGTGGA	(KanMX4-Mu) GTGGAgaagactcga	chr14:661338-661342
G41	ttcactcattCTGGT	(KanMX4-Mu) CTGGTcatttcttcg	chr15:720163-720167
G42	ctagcgctttACGGA	(KanMX4-Mu) ACGGAagacaatgta	2μ:2838-2842
G43	ggtaataggcCCGTG	(KanMX4-Mu) CCGTGcggttccgtc	chr15:836789-836793
G44	gtggtgccctTCCGT	(KanMX4-Mu) TCCGTcaattccttt	chr12:456583-456587
G45	ttcgctgctcACCAA	(KanMX4-Mu) ACCAAtggaatcgca	chr12:458164-458168
G46	aatattatctTCTGT	(KanMX4-Mu) TCTGTcattgttact	chr10:135624-135628
G47	gtatgtacccACCGA	(KanMX4-Mu) ACCGAtgtagcagta	chr15:829039-829043
G48	gttgatggtaCCTTG	(KanMX4-Mu) CCTTGacaccagcca	chr6:44321-44325
G49	tacattgtctTCCGT	(KanMX4-Mu) TCCGTaaagcgctag	2μ:2838-2842
G50	ccgtggaagcCTCGC	(KanMX4-Mu) CTCGCccgatgagtt	chr10:526881-526885
G51	tttcttttccTCCGC	(KanMX4-Mu) TCCGCttattgatat	chr12:455126-455130
G52	gctgcgtctgACCAA	(KanMX4-Mu) ACCAAggccctcact	chr12:453213-453217
G53	tactgttgaaCCGGG	(KanMX4-Mu) CCGGGtcgtacaact	chr14:736161-736165
G54	caaatgtatcAGCAG	(KanMX4-Mu) AGCAGatgtacttcc	chr14:566860-566864
G55	agtttccgctATAAA	(KanMX4-Mu) ATAAAtaatggcagc	chr10:161496-161500
G56	aaaggaattgCTAGG	(KanMX4-Mu) CTAGGggcattactc	chr12:912615-912619
G57	aaaaataattACTCT	(KanMX4-Mu) ACTCTaacatttctt	chr16:120160-120164
G58	tgtttatatgATGAC	(KanMX4-Mu) ATGACgattttccca	chr11:306835-306839
G59	ttgtgtatttTTGAT	(KanMX4-Mu) TTGATtgaaaatgat	chr4:600461-600465
G60	tatgataatcAAGGC	(KanMX4-Mu) AAGGCataattgact	chr2:429112-429116
G63	cagcattaaaACGGC	(KanMX4-Mu) ACGGCagcaaagccc	chr16:826635-826639
G64	ttgacatgtgATCTG	(KanMX4-Mu) ATCGTcacagatttt	2μ:5268-5272
G65	tcagctctcaGCAGA	(KanMX4-Mu) GCAGAgaaaaaattt	chr2:117272-117276
G66	tgctaggtgtGTCTG	(KanMX4-Mu) GTCTGtttatgcatt	chr14:331432-331436
G67	caattgaggtTTGAA	(KanMX4-Mu) TTGAAattgctggcc	chr.12:455361-455365
G67	aatcatgcatTGCAT	(KanMX4-Mu) TGCATaatgtggtat	2μ:2196-2200
G70	acgatcttacGTCGG	(KanMX4-Mu) GTCGGctatctcacc	chr3:77666-77670
G71	ttgtatttaaACTGG	(KanMX4-Mu) ACTGGagtgatttat	2μ A:5800-5804
G74	tgcatatttgCCTGC	(KanMX4-Mu) CCTGCgaaaaaaagt	chr5:436799-436803
G75	tcgttgaataATGGA	(KanMX4-Mu) ATGGAaaatatgaaa	chr10: 187594-187598

Table 3A (Continued)

G76	ctttcccagaACCAG(KanMX4-Mu)ACCAGggaaactgtt	chr14: 537718-537722
G77	cctctgcatcCCAAC(KanMX4-Mu)CCAACaccagcgata	chr4:955105-955109
G78	atctgtaaacTCGCT(KanMX4-Mu)TCGCTtgtgacgatg	chr4: 480341-480435
G79	tcctgcctaaACAGG(KanMX4-Mu)ACAGGaagacaaagc	chr14: 547141-547145
G80	tagaaaaaacCACAA(KanMX4-Mu)CACAAcaacactatg	chr10: 111531-111535
G81	ttttggctcgTCCGG(KanMX4-Mu)TCCGGatgatgcgaa	chr.16:641397-641401
G83	tgtggctaccGCCCG(KanMX4-Mu)GCCCGtgattcgggc	chr4:1433822-1433826
G84	ggcatagtgcGTGTT(KanMX4-Mu)GTGTTtatgcttaaa	2μ:541-545
G85	aaaatgcaacGCGAG(KanMX4-Mu)GCGAGagcgctaatt	2μ:3134-3138
G87	gaacagttccACGCC(KanMX4-Mu)ACGCCtgatatgagg	chr11: 60765-60769
G88	agcgcgactgCCCGA(KanMX4-Mu)CCCGAagaaggacgc	chr4:1056229-1056233
G90	aaaaggttcaGTAGA(KanMX4-Mu)GTAGAaacataaaat	chr11:430889-430893
G94	ccacaaggacGCCTT(KanMX4-Mu)GCCTTattcgtatcc	chr12: 451993-451997
G96	cagaatccatGCTAG(KanMX4-Mu)GCTAGaacgcggtga	chr12:452043-452047
G97	cagctgctacCCAGG(KanMX4-Mu)CCAGGgattgccacg	chr2:415433-415437
G98	ctagccgttcATCAA(KanMX4-Mu)ATCAAtcatgtcaaa	chr4:539356-539360
G99	caaaaaagtcTAGAG(KanMX4-Mu)TAGAGgaaaaaaacg	chr13:406197-406201
G100	ttgtcaaagtACCGA(KanMX4-Mu)ACCGAtcatgacaat	chr5:258808-258812
G101	gtaacatcttGGGCG(KanMX4-Mu)GGGCGtttgcaacac	chr16:135372-135376
G102	actgcctttgCTGAG(KanMX4-Mu)CTGAGctggatcaat	2μ:2524-2528
G103	aatgtaaaagGCAAG(KanMX4-Mu)GCAAGaaaacatgta	chr4: 1011940-1011944
G104	gcctgaattgTAGAT(KanMX4-Mu)TAGATattagataag	chr15: 770712-770716
G105	gtttgacattGTGAA(KanMX4-Mu)GTGAAgagacataga	chr12:452744-452748
G106	tgtcatctacATCAT(KanMX4-Mu)ATCATcggtattatt	chr4: 1160847-1160851
G107	cttgttcctaGTGGC(KanMX4-Mu)GTGGCgctaatggga	chr4:464844-464848
G108	agggccctcaGTGAT(KanMX4-Mu)GTGATggtgttttgt	2μ B:4396-4400
G109	ggtattttcaTTGGT(KanMX4-Mu)TTGGTtgtaaaatcg	chr12:582690-582694
G110	caatctaaccACCAT(KanMX4-Mu)ACCATgttggctcac	chr15:75760-75764
G111	cgaaaaatgcACCGG(KanMX4-Mu)ACCGGccgcgcatta	2μ:5427-5431
G113	ttacgatctgCTGAG(KanMX4-Mu)CTGAGattaagcctt	chr12:451812-451816
G114	aaatcgagcaATCAC(KanMX4-Mu)GTGATtgctcgattt	2μ:2126-2130
G116	ccgacaaaccCCCCC(KanMX4-Mu)CCCCCcatttatata	chr15:1039713-1039717
G117	caataagatgTGGGG(KanMX4-Mu)TGGGGattagtttcg	chr13:895900-895904
G118	gtttaacgctTCCTG(KanMX4-Mu)TCCTGggaactgcag	chr16:30277-30281
G120	atgaatactcCTCCC(KanMX4-Mu)CTCCCttgctgttgg	chr14:175588-175592
G121	aatcacaatgGCGGC(KanMX4-Mu)GCGGCcatcgaccct	chr12:1030933-1030937
G122	gagcaccacgATCGT(KanMX4-Mu)ATCGTtcggtgtact	chr13:67812-67816
G123	aaaagcattcTGCAG(KanMX4-Mu)TGCAGtaattagccg	chr15:638922-638926
G124	gtgattctccATGGG(KanMX4-Mu)ATGGGtggtttcgct	chr14:333823-333827
G125	gctggtccagACCAC(KanMX4-Mu)ACCACaaaaggatgc	chr13:540587-540591
G126	acttcgacttCGGGT(KanMX4-Mu)CGGGTaaaatactct	chr12:328174-328178
G127	tgacattaatCCTAC(KanMX4-Mu)CCTACgtgacttaca	chr5:291453-291457
G128	tttatatccgGTGGT(KanMX4-Mu)GTGGTtgcgataagg	chr5:317469-317473
G129	ctgatgtgcgGTGGT(KanMX4-Mu)GTGGGccttggactt	chr5:336404-336408
G130	gttgaactacTACGG(KanMX4-Mu)TACGGttaagggtgc	chr16:40318-40322
G131	cctatactctACCGT(KanMX4-Mu)ACCGTcagggttgat	chr12:453842-453846
G132	aactagcaaaATGGA(KanMX4-Mu)ATGGAaacaaaaaaa	chr2:692001-692005
G133	ttgactcaacACGGG(KanMX4-Mu)ACGGGgaaactcacc	chr12:456534-456538
G134	cattgtgaccCTGGC(KanMX4-Mu)CTGGCaaatttgcaa	chr12:651930-651934
G135	atacagctcaCTGTT(KanMX4-Mu)CTGTTcacgtcgcac	2μ B:4039-4043
G136	tcagatttttCCCAG(KanMX4-Mu)CCCAGtatggctttg	chr7:976865-976869
G137	tttaacgtggGCGAA(KanMX4-Mu)GCGAAgaagaaggaa	chr11:327312-327316
G138	ccattccataTCTGT(KanMX4-Mu)TCTGTtaagtataca	chr12:460247-460251
G140	ctttgtgcgcTCTAT(KanMX4-Mu)TCTATaatgcagtct	2μ:3318-3322
G150	aattggtacaGTATG(KanMX4-Mu)GTATGctcaaaaata	chr12:492584-492588
T1	ttgtagcttcCACAA(Mu-KanMX4-p15A-Mu)CACAAgatgttggct	chr12:645643-645647
T2	tcttattctcCTGTT(Mu-KanMX4-p15A-Mu)CTGTTgccttcgtac	chr5:7908-7912
T3	cggttgtataTGCAT(Mu-KanMX4-p15A-Mu)TGCATtgtacgtgcg	chr5:402750-402754
T4	ttttaataagGCAAT(Mu-KanMX4-p15A-Mu)GCAATaatattaggt	chr10:538071-538075

Table 3A (Continued)

T5	tatcacttacTCGAA(Mu-KanMX4-p15A-Mu)TCGAAcgttgacatt	chr12:864259-864263
T6	aaagacatctACCGT(Mu-KanMX4-p15A-Mu)ACCGTgaaggtgccg	chr7:999996-1000000
T7	catattactgCCCGC(Mu-KanMX4-p15A-Mu)CCCGCgtaatccaat	chr15:304883-304887
T8	gtgttagtgaATGCC(Mu-KanMX4-p15A-Mu)ATGCCtcaaactctt	chr10:304087-304091

Target site duplication is typed in capital letters.
*Chromosome and the coordinates of the duplicated sequence.

Table 3B. Transposon integration sites and target site duplications in *Saccharomyces cerevisiae* haploid strain FY-3.

	←seqmx4	seqA→	Location*
G1	aaagagaaaaATAAG	(KanMX4-Mu) ATAAGaaaatcttct	chr3:38982-38986
G2	cctttttttcGTGGG	(KanMX4-Mu) GTGGGaaccgcttta	2μ: A:4372-4376
G3	atccacctttGCTGC	(KanMX4-Mu) (GCTGCttttccttaa)	2μ:5349-5353
G4	tacattcctcCTCAT	(KanMX4-Mu) CTCATttgaccgagg	chr16: 837554-837558
G5	gatttatcatGCAGT	(KanMX4-Mu) GCAGTaatactaata	chr4: 3069-3073
G6	gaattttaaGAGAtc	(KanMX4-Mu) GAtcAAgtcttgtga	chr15: 144910-144915
G7	gttcgatgctGTGCG	(KanMX4-Mu) GTGCGggacttctac	chr1: 191076-191080
G8	cttcacggtaACGTA	(KanMX4-Mu) ACGTAactgaatgtg	chr12:453541-453545
G9	caaggagcagAGGGC	(KanMX4-Mu) AGGGCacaaaacacc	chr12:454727-424731
G10	tcaataaacaGCCGA	(KanMX4-Mu) GCCGAcatacatccc	2μ:5123-5127
G11	gcgagatgagGTGAA	(KanMX4-Mu) GTGAAaagaaactta	chr7: 284048-284052
G12	taaatttcatCCGGA	(KanMX4-Mu) CCGGAagaaaaatga	chr11:489457-489461
G13	agaaaagtacAATTc	(KanMX4-Mu) gATcAaggttacggc	chr4: 56735-56740
G14	actgtcttttCCGGT	(KanMX4-Mu) CCGGTcattccaaca	chr11: 428648-428652
G15	atacacgctcATCAG	(KanMX4-Mu) ATCAGacaccacaaa	chr12:453989-453993
G16	atagtatttcCTAGT	(KanMX4-Mu) CTAGTgatctcggcg	chr15: 989676-989680
G17	ttcctattctCTAGA	(KanMX4-Mu) CTAGAaagtatagga	2μ:704-708
G28	ttataaggttGTTTC	(KanMX4-Mu) gaGTTTCatatgtgttt	chr15:854340-854344
G37	ttcgagagtgCCATT	(KanMX4-Mu) CCATTgtaccagact	chr8:489155-489159
G38	atggatggcgCTCAA	(KanMX4-Mu) CTCAAgcgtgttacc	chr12:453865-453869
G39	tccaaatgtaTTGTG	(KanMX4-Mu) TTGTGagatgaaaat	chr15:834888-834892
G40	atgattatttCACGG	(KanMX4-Mu) CACGGatttcattag	chr13:97657-97661
G42	atggaaaactAGCGC	(KanMX4-Mu) AGCGCataattttgt	chr4:437081-437085
G43	gagaatcttgTCTTG	(KanMX4-Mu) TCTTGatgtaacaaa	chr7:190765-190769
G44	tagcaaacgTAAGTCTtc	(KanMX4-Mu) gAAGTCTAAaggttg	chr12:459205-459213
G45	ttgccgcgaaGCTAC	(KanMX4-Mu) GCTACcatccgctgg	chr12:452091-452095
G46	gtagctctttTCCAT	(KanMX4-Mu) TCCATggatggacga	chr12:645493-645497
G47	atgttcattcTCTGT	(KanMX4-Mu) TCTGTagcagtaaga	chr10:337762-337766
G48	aatcgtaaccATAAA	(KanMX4-Mu) ATAAAtataagttcc	chr2:806825-806829
G49	ccttcctgctGTGGG	(KanMX4-Mu) GTGGGcagagagcga	chr7:739278-739278
G50	tcttagggttATTGG	(KanMX4-Mu) ATTGGtagggttttg	chr9:382384-382388
G51	agttaacttcCCCGG	(KanMX4-Mu) CCCGGtgttcagtat	chr12:1025073-1025077
G52	atgtgtcattGAGGG	(KanMX4-Mu) GAGGGaaaatgtaat	chr7:798084-798088
G53	ggttaacttgCTCGC	(KanMX4-Mu) CTCGCcatatatatc	chr2:657457-657461
G54	caaaaaaagaTGGAG	(KanMX4-Mu) TGGAGtacagtacgc	chr2:466108-466112
G55	gatatttacgCTTAT	(KanMX4-Mu) CTTATcaatctctgg	chr2:80588-80592
G56	gccgtggtttCCGGA	(KanMX4-Mu) CCGGAgaaagacgaa	chr13:347229-347233
G57	tttctggaatTAGGG	(KanMX4-Mu) TAGGGtgacagaatg	chr4:722468-722472
G58	attactttatTTGGC	(KanMX4-Mu) TTGGCtaaagatcct	chr4:600407-600411
G59	cgttatcataTTGAT	(KanMX4-Mu) TTGAtattgcttatt	chr15:696010-696013
G60	ggcaaactatCTCAC	(KanMX4-Mu) CTCACcagaggtctg	chr10:117057-117061
G61	ctaatagtgcATGAT	(KanMX4-Mu) ATGATtatatatcaa	chr7:853604-853608
G62	agaaattctcCTTGG	(KanMX4-Mu) CTTGGgattagataa	chr5:137549-137553
G63	tcccgcactgGTGAT	(KanMX4-Mu) GTGATacctacaccc	chr12:213298-213302
G64	atcattcattGCCGG	(KanMX4-Mu) GCCGGaaaaagaaag	chr12:370966-370970
G65	ctcacgctctGCGAT	(KanMX4-Mu) GCGATtaacagctca	chr10:404834-404888

Target site duplication is typed in capital letters.
*Chromosome and the coordinates of the duplicated sequence.

Table 4. Nucleotide consensus of the sequenced duplicated pentamers. (Percentage)

FY1679 (n=121):

Nucleotide	1	2	3	4	5
A	34 (28)	10 (8)	13 (11)	47 (39)	27 (22)
C	31 (26)	58 (48)	45 (37)	8 (7)	27 (22)
G	28 (23)	11 (9)	49 (41)	53 (44)	36 (30)
T	28 (23)	42 (35)	14 (12)	13 (11)	31 (26)
Consensus:	N	C/T	C/G	A/G	N

FY-3 (n=42):

Nucleotide	1	2	3	4	5
A	8 (19)	3 (7)	6 (14)	15 (36)	8 (19)
C	14 (33)	15 (36)	11 (26)	1 (2)	7(17)
G	12 (28)	3 (7)	18 (42)	22 (51)	15 (35)
T	8 (19)	21 (50)	7 (18)	4 (10)	12 (29)
Consensus:	N	C/T	C/G	A/G	N

FY1679 + FY-3 (n=163):

Nucleotide	1	2	3	4	5
A	42 (26)	13 (8)	19 (12)	62 (38)	35 (21)
C	45 (28)	73 (45)	56 (34)	9 (6)	34 (21)
G	40 (25)	14 (9)	67 (41)	75 (46)	51 (31)
T	36 (22)	63 (39)	21 (13)	17 (10)	43 (26)
Consensus:	N N	C/T		C/G	A/G

Table 5. Transposon integration sites and target site duplications in HeLa cells.

Clone		Location*
RGC16	aggaggaagaACCAG(Mu/LoxP-Kan/Neo)ACCAGgcacatgctg	chr8 : 128251032-128251036
RGC26	ttaaatgaacTTCAG(Mu/LoxP-Kan/Neo)TTCAGgaaaataatg	chr12 : 15381980-15381984
RGC35	ccaatgagtcACCAG(Mu/LoxP-Kan/Neo)ACCAGaactgaacaa	chr2 : 180174041-180174045

Target site duplication is typed in capital letters.

*Chromosome and the coordinates of the duplicated sequence.

REFERENCES

Allet, B. (1979). Mu insertion duplicates a 5 base pair sequence at the host inserted site. Cell *1,* 123-129.

Ausubel, F.M., Brent, R., Kingston, R.E., Moore, D.D., Seidman, J.G., Smith, J.A., and Struhl, K. (1989). Current protocols in molecular biology. John Wiley & Sons, New York, N.Y.

Behrens, R., Hayles, J., and Nurse, P. (2000). Fission yeast retrotransposon Tf1 integration is targeted to 5' ends of open reading frames. Nucleic Acids Res. *23,* 4709-4716.

Butterfield, Y.S., Marra, M.A., Asano, J.K., Chan, S.Y., Guin, R., Krzywinski, M.I., Lee, S.S., MacDonald, K.W., Mathewson, C.A., and Olson, T.E. *et al.* (2002). An efficient strategy for large-scale high-throughput transposon-mediated sequencing of cDNA clones. Nucleic Acids Res. *11,* 2460-2468.

Chaconas, G., Lavoie, B.D., and Watson, M.A. (1996). DNA transposition: jumping gene machine, some assembly required. Curr. Biol. *7,* 817-820.

Haapa, S., Suomalainen, S., Eerikainen, S., Airaksinen, M., Paulin, L., and Savilahti, H. (1999a). An efficient DNA sequencing strategy based on the bacteriophage mu in vitro DNA transposition reaction. Genome Res. *3,* 308-315.

Haapa, S., Taira, S., Heikkinen, E., and Savilahti, H. (1999b). An efficient and accurate integration of mini-Mu transposons in vitro: a general methodology for functional genetic analysis and molecular biology applications. Nucleic Acids Res. *13,* 2777-2784.

Haapa-Paananen, S., Rita, H., and Savilahti, H. (2002). DNA transposition of bacteriophage Mu. A quantitative analysis of target site selection in vitro. J. Biol. Chem. *4,* 2843-2851.

Ji, H., Moore, D.P., Blomberg, M.A., Braiterman, L.T., Voytas, D.F., Natsoulis, G., and Boeke, J.D. (1993). Hotspots for unselected Ty1 transposition events on yeast chromosome III are near tRNA genes and LTR sequences. Cell *5,* 1007-1018.

Kahmann, R., and Kamp, D. (1979). Nucleotide sequences of the attachment sites of bacteriophage Mu DNA. Nature *5719,* 247-250.

Kekarainen, T., Savilahti, H., and Valkonen, J.P. (2002). Functional genomics on potato virus A: virus genome-wide map of sites essential for virus propagation. Genome Res. *4,* 584-594.

Lamberg, A., Nieminen, S., Qiao, M., and Savilahti, H. (2002). Efficient insertion mutagenesis strategy for bacterial genomes involving electroporation of in vitro-assembled DNA transposition complexes of bacteriophage mu. Appl. Environ. Microbiol. *2,* 705-712.

Laurent, L.C., Olsen, M.N., Crowley, R.A., Savilahti, H., and Brown, P.O. (2000). Functional characterization of the human immunodeficiency virus type 1 genome by genetic footprinting. J. Virol. *6,* 2760-2769.

Livingston, D.M., and Hahne, S. (1979). Isolation of a condensed, intracellular form of the 2-micrometer DNA plasmid of Saccharomyces cerevisiae. Proc. Natl. Acad. Sci. U. S. A. *8,* 3727-3731.

Miller, S.A., Dykes, D.D., and Polesky, H.F. (1988). A simple salting out procedure for extracting DNA from human nucleated cells. Nucleic Acids Res. *3,* 1215.

Mizuuchi, K. (1992). Transpositional recombination: mechanistic insights from studies of mu and other elements. Annu. Rev. Biochem. 1011-1051.

Mizuuchi, M., and Mizuuchi, K. (1993). Target site selection in transposition of phage Mu. Cold Spring Harb. Symp. Quant. Biol. 515-523.

Nelson, R.G., and Fangman, W.L. (1979). Nucleosome organization of the yeast 2-micrometer DNA plasmid: a eukaryotic minichromosome. Proc. Natl. Acad. Sci. U. S. A. *12,* 6515-6519.

Raz, E., van Luenen, H.G.A.M., Schaerringer, B., Plasterk, R.H.A., and Driever, W. (1997). Transposition of the nematode Caenohabditis elegans Tc3 element in the zebrafish Danio rerio. Current Biology 8, 82-88.

Rose, R.E. (1988). The nucleotide sequence of pACYC184. Nucleic Acids Res. *1,* 355.

Rubin, G.M., and Spradling, A.C. (1982). Genetic transformation of Drosophila with transposable element vectors. Science 218, 348-353.

Sambrook, J., Fritch, E.F., and Maniatis, T. (1989). Molecular cloning: a laboratory manual, 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.

Sands, A.T., and Hasty, E.P. eds. (1997). Mouse mutagenesis, Lexicon Genetics Incorporated: Houston, Texas.

Savilahti, H., and Mizuuchi, K. (1996). Mu transpositional recombination: donor DNA cleavage and strand transfer in trans by the Mu transposase. Cell *2,* 271-280.

Savilahti, H., Rice, P.A. and Mizuuchi, K. (1995) The phage Mu transpososome core: DNA requirements for assembly and function. EMBO J. 14, 4893-4903

Sherman, F., Fink, G.R. and Lawrence, C.W. (1981) Methods in yeast genetics. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York

Taira, S., Tuimala, J., Roine, E., Nurmiaho-Lassila, E.L., Savilahti, H., and Romantschuk, M. (1999). Mutational analysis of the Pseudomonas syringae pv. tomato hrpA gene encoding Hrp pilus subunit. Mol. Microbiol. *4,* 737-744.

Taketo, M., Jazwinski, S.M., and Edelman, G.M. (1980). Association of the 2-micron DNA plasmid with yeast folded chromosomes. Proc. Natl. Acad. Sci. U. S. A. *6,* 3144-3148.

Vilen, H., Eerikainen, S., Tornberg, J., Airaksinen, M.S., and Savilahti, H. (2001). Construction of gene-targeting vectors: a rapid Mu in vitro DNA transposition-based

strategy generating null, potentially hypomorphic, and conditional alleles. Transgenic Res. *1*, 69-80.

Vilen, H., Aalto, J-M., Kassinen, A., Paulin, L., and Savilahti, H. (2003). A direct transposon insertion tool for modification and functional analysis of viral genomes. J.Virol. 77, 123-134.

Wach, A., Brachat, A., Pohlmann, R., and Philippsen, P. (1994). New heterologous modules for classical or PCR-based gene disruptions in Saccharomyces cerevisiae. Yeast *13*, 1793-1808.

Winston, F., Dollard, C., and Ricupero-Hovasse, S.L. (1995). Construction of a set of convenient Saccharomyces cerevisiae strains that are isogenic to S288C. Yeast *1*, 53-55.

We claim:

1. A method for incorporating nucleic acid segments into cellular nucleic acid of an isolated eukaryotic target cell, the method comprising the step of:

delivering into the eukaryotic target cell an *in vitro* assembled Mu transposition complex that comprises (i) MuA transposases and (ii) a transposon segment that comprises a pair of Mu end sequences recognised and bound by MuA transposase and an insert sequence between said Mu end sequences, under conditions that allow integration of the transposon segment into the cellular nucleic acid.

2. The method according to claim 1, wherein said Mu transposition complex is delivered into the target cell by electroporation.

3. The method according to claim 1, wherein the nucleic acid segment is incorporated to a random or almost random position of the cellular nucleic acid of the target cell.

4. The method according to claim 1, wherein the nucleic acid segment is incorporated to a targeted position of the cellular nucleic acid of the target cell.

5. The method according to claim 1, wherein the target cell is human, animal, plant, fungi or yeast cell

6. The method according to claim 5, wherein said animal cell is a mouse cell.

7. The method according to claim 1, wherein said insert sequence comprises a marker, which is selectable in eukaryotic cells.

8. The method according to claim 1, wherein a concentrated fraction of Mu transposition complexes are delivered into the target cell.

9. The method according to claim 1 further comprising the step of incubating the target cells under conditions that promote transposition into the cellular nucleic acid.

10. A method for forming an insertion mutant library from a pool of eukaryotic target cells, the method comprising the steps of:

a) delivering into the eukaryotic target cell an *in vitro* assembled Mu transposition complex that comprises (i) MuA transposases and (ii) a transposon segment that comprises a pair of Mu end sequences recognised and bound by MuA transposase and an insert sequence with a selectable marker between said Mu end sequences, under conditions that allow integration of the transposon segment into the cellular nucleic acid; and

b) screening for cells that comprise the selectable marker.

11. A kit for incorporating nucleic acid segments into cellular nucleic acid of a eukaryotic target cell comprising a concentrated fraction of Mu transposition complexes with a transposon segment that comprises a marker, which is selectable in eukaryotic cells.

SEQUENCE LISTING

<110> Finnzymes Oy

<120> Method for delivering nucleic acid into eukaryotic genomes

<140>
<141>

<160> 21

<170> PatentIn Ver. 2.1

<210> 1
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Oligonucleotide primer

<400> 1
gctctccccg tggaggtaat 20

<210> 2
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Oligonucleotide primer

<400> 2
ttccgtcaca ggtatttatt cggt 24

<210> 3
<211> 17
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Oligonucleotide primer

<400> 3
atcagcggcc gcgatcc 17

<210> 4
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Oligonucleotide primer

<400> 4
ggacgaggca agctaaacag 20

<210> 5
<211> 33
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:
Oligonucleotide primer

<400> 5
ctaataccac tcacataggg cggccgcccg ggc 33

<210> 6
<211> 13
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:
Oligonucleotide primer

<400> 6
gatcgcccgg gcg 13

<210> 7
<211> 13
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:
Oligonucleotide primer

<400> 7
ctaggcccgg gcg 13

<210> 8
<211> 13
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:
Oligonucleotide primer

<400> 8
aattgcccgg gcg 13

<210> 9
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:
Oligonucleotide primer

<400> 9
ctaataccac tcacataggg 20

<210> 10
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:
Oligonucleotide primer

<400> 10
gggcggccgc ccgggcgatc 20

<210> 11
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:
Oligonucleotide primer

<400> 11
gggcggccgc ccgggcctag 20

<210> 12
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:
Oligonucleotide primer

<400> 12
gggcggccgc ccgggcaatt 20

<210> 13
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:
Oligonucleotide primer

<400> 13
ctgtcgattc gatactaacg 20

<210> 14
<211> 27
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:
Oligonucleotide primer

<400> 14
ctctagatga tcagcggccg cgatccg 27

<210> 15
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:
Oligonucleotide primer

<400> 15
tgtcaaggag ggtattctgg 20

<210> 16
<211> 22
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:
Oligonucleotide primer

<400> 16
ggtgacccgg cggggacgag gc 22

<210> 17
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:
Oligonucleotide primer

<400> 17
gatccgtttt cgcatttatc gtg 23

<210> 18
<211> 28
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:
Oligonucleotide primer

<400> 18
ggccgcatcg ataagcttgg gctgcagg 28

<210> 19
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:
Oligonucleotide primer

<400> 19
acattgggtg gaaacattcc 20

<210> 20
<211> 18
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:
Oligonucleotide primer

<400> 20
ccaagttcgg gtgaaggc 18

<210> 21
<211> 22
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:
Oligonucleotide primer

<400> 21
ccccgggcga gtctagggcc gc 22

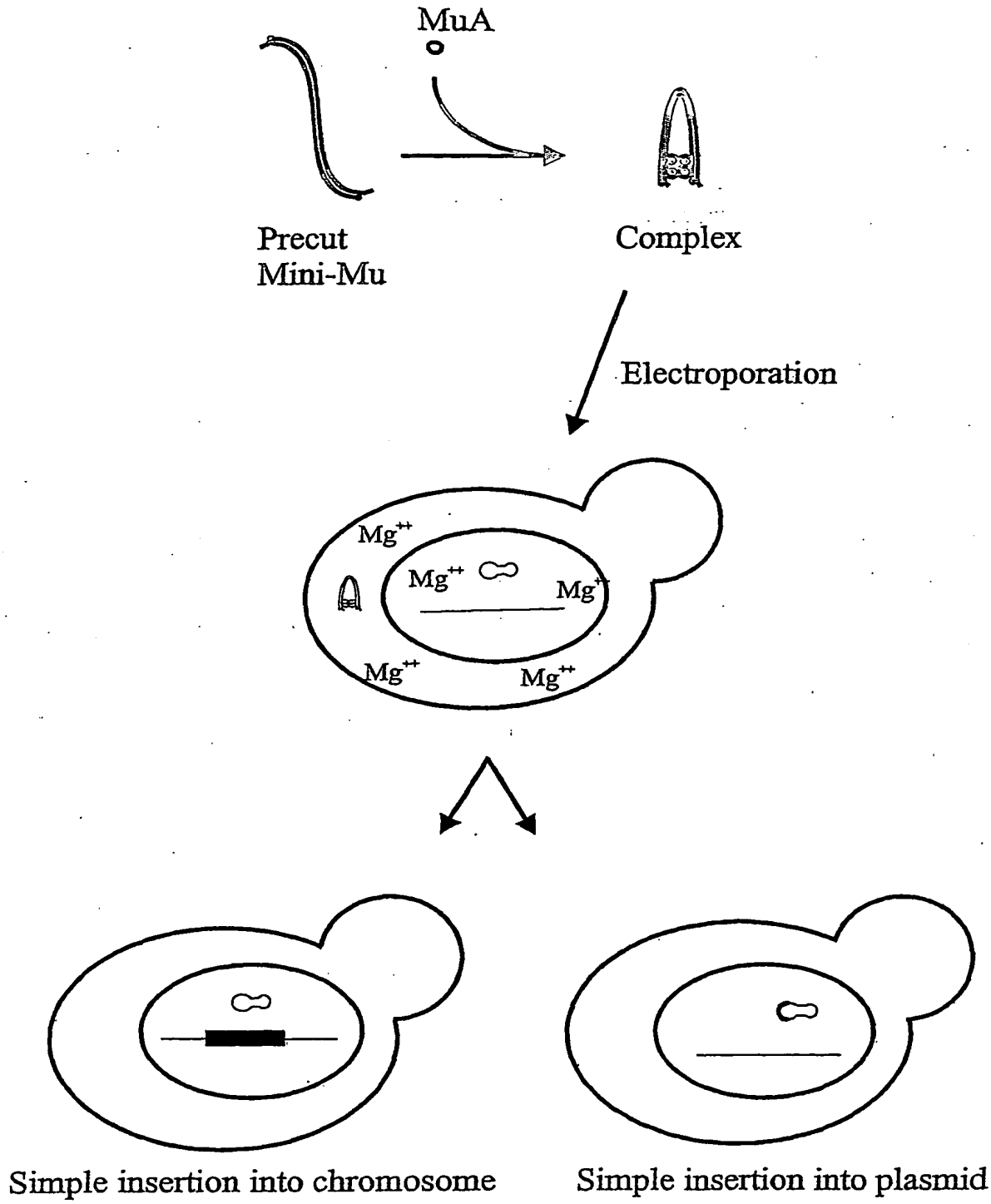

MuA
Precut
Mini-Mu
Complex
Electroporation
Mg++
Mg++
Mg++
Mg++
Mg++
Simple insertion into chromosome
Simple insertion into plasmid

Fig. 1

A

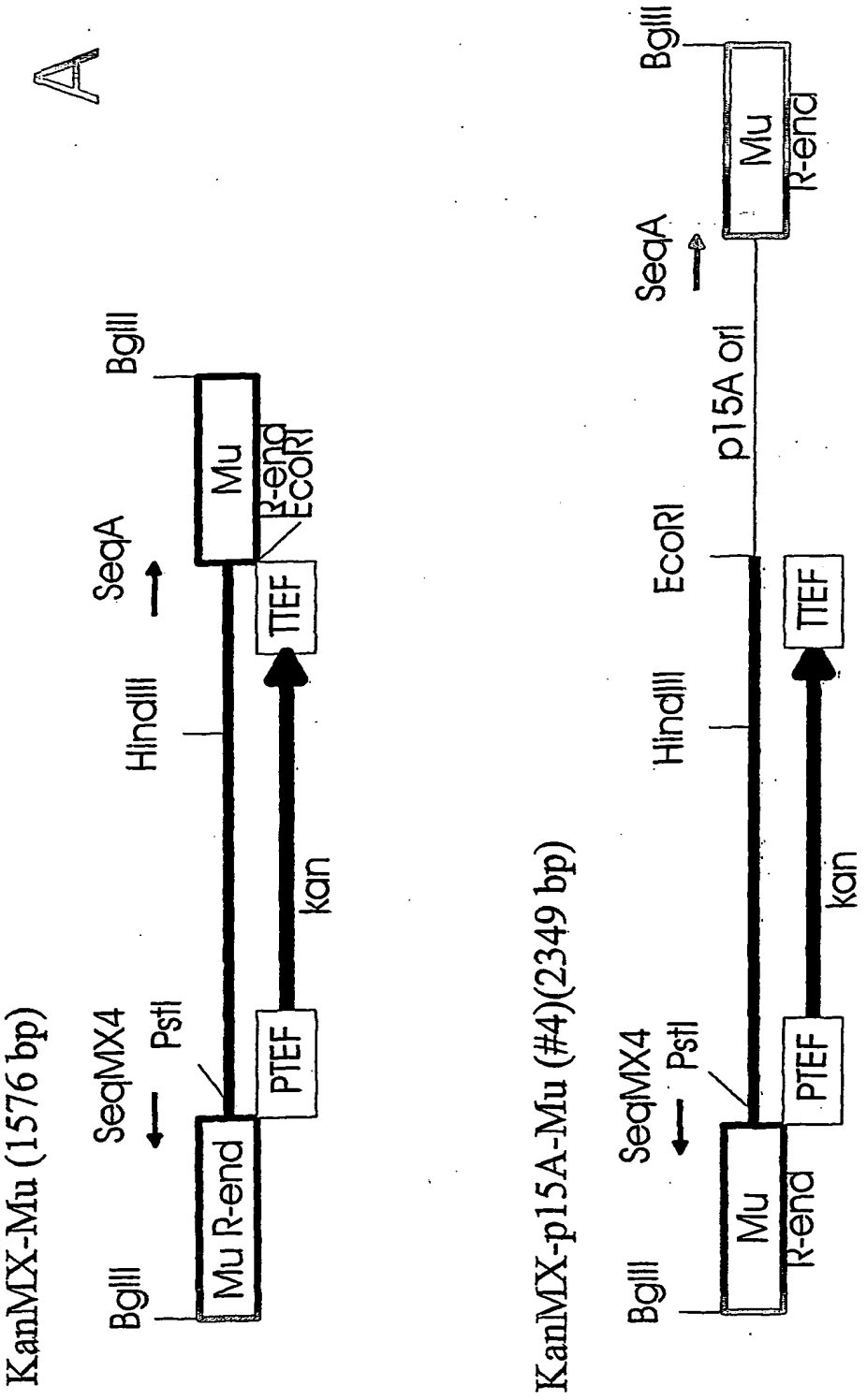

KanMX-Mu (1576 bp)
BglII
Mu R-end
SeqMX4
PstI
PTEF
kan
HindIII
SeqA
BglII
Mu
R-end
EcoRI
TTEF
KanMX-p15A-Mu (#4)(2349 bp)
BglII
Mu
R-end
SeqMX4
PstI
PTEF
kan
HindIII
EcoRI
p15A ori
SeqA
BglII
Mu
R-end
TTEF

Fig. 2A

B

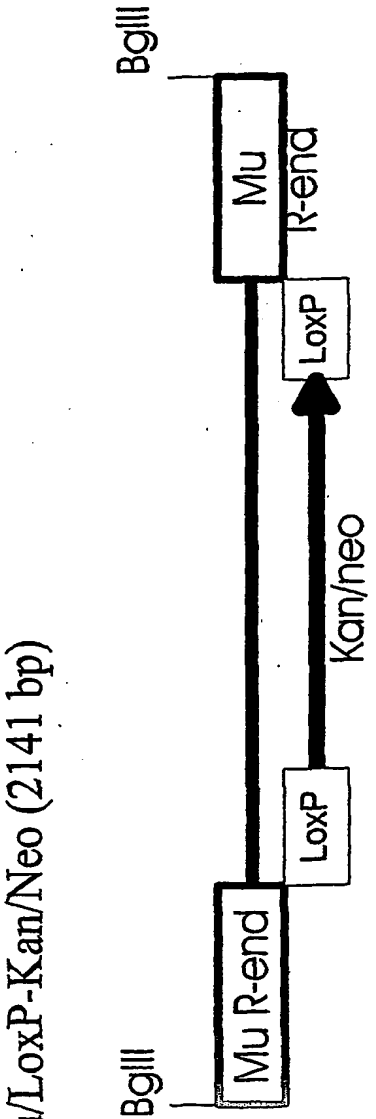


Fig. 2B

Mu transposition complex formation

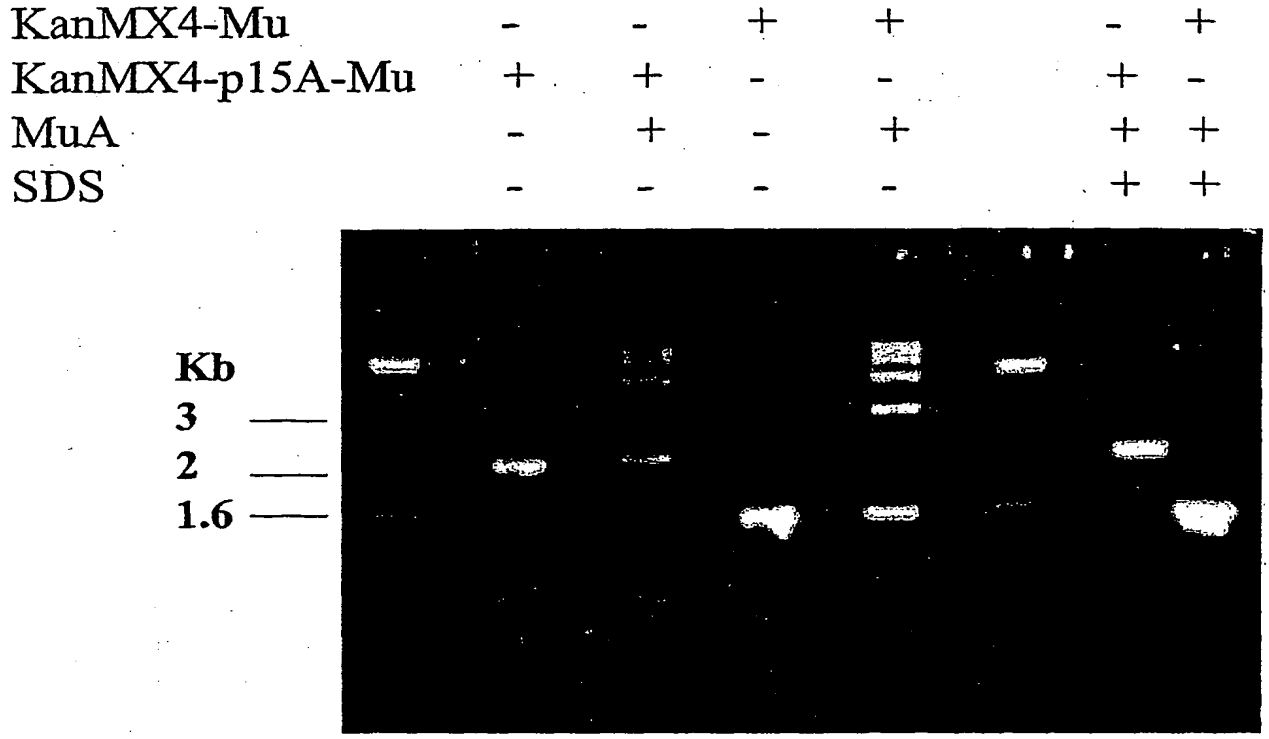
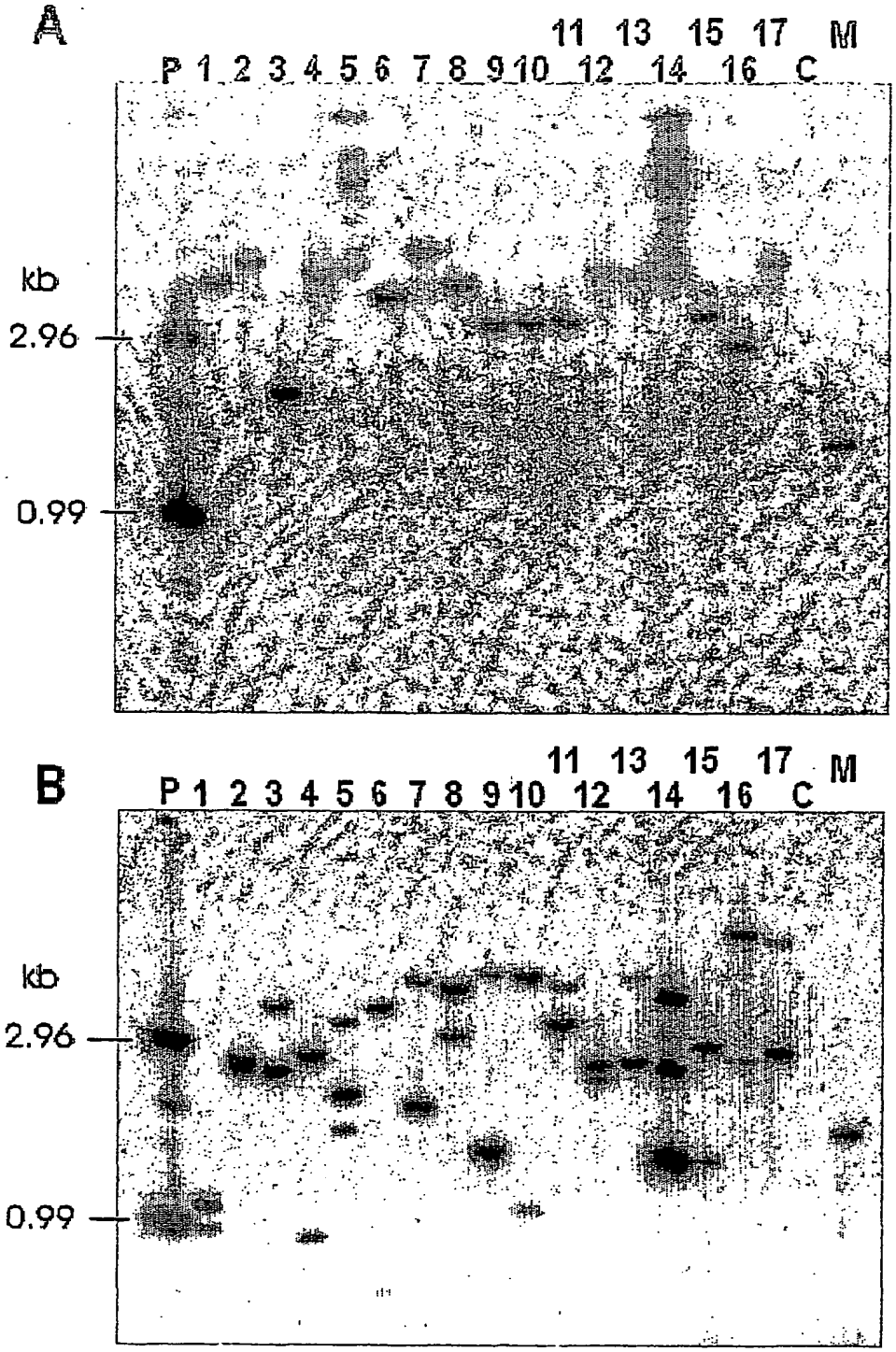


Fig. 3



Figs. 4A and 4B

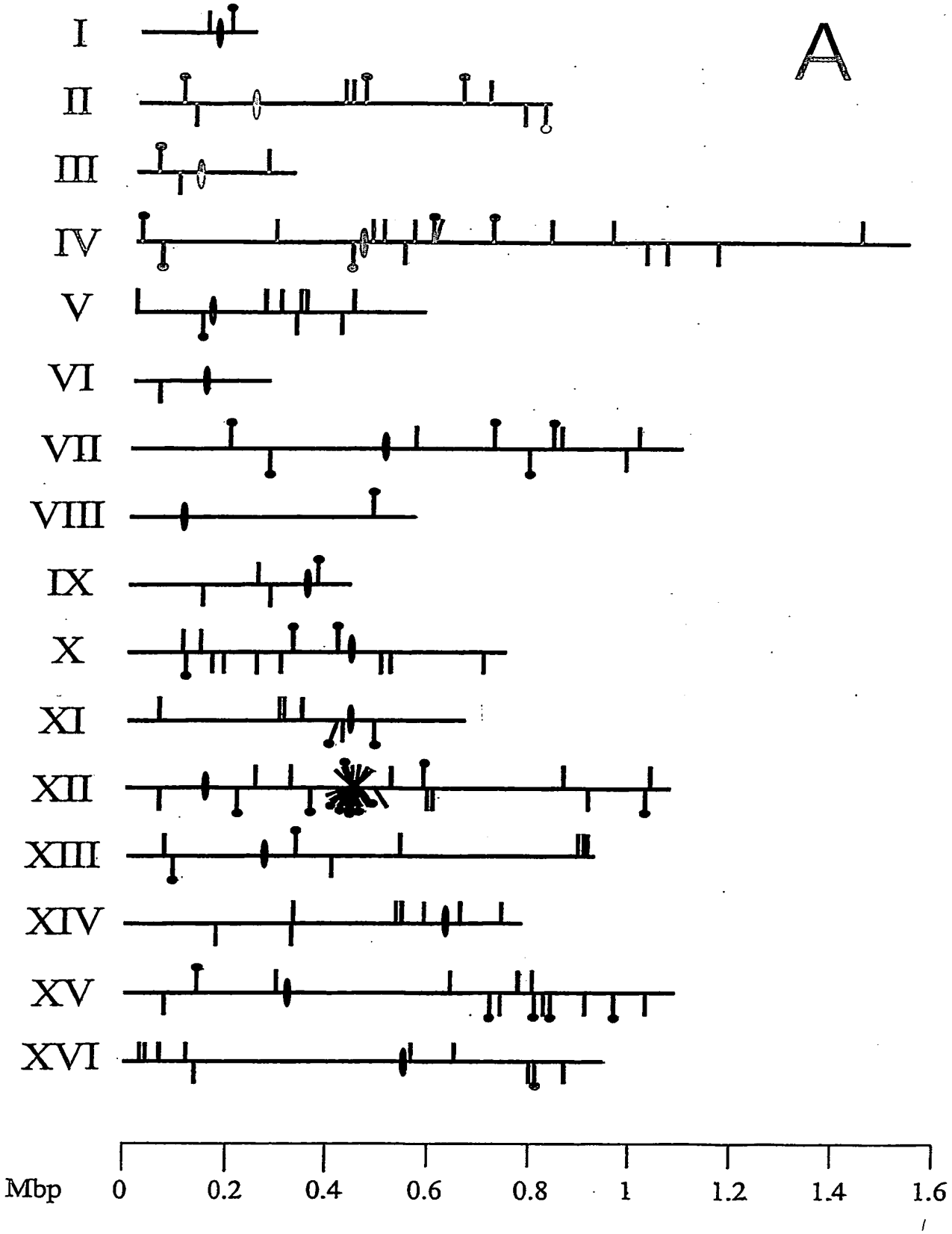


Fig. 5A

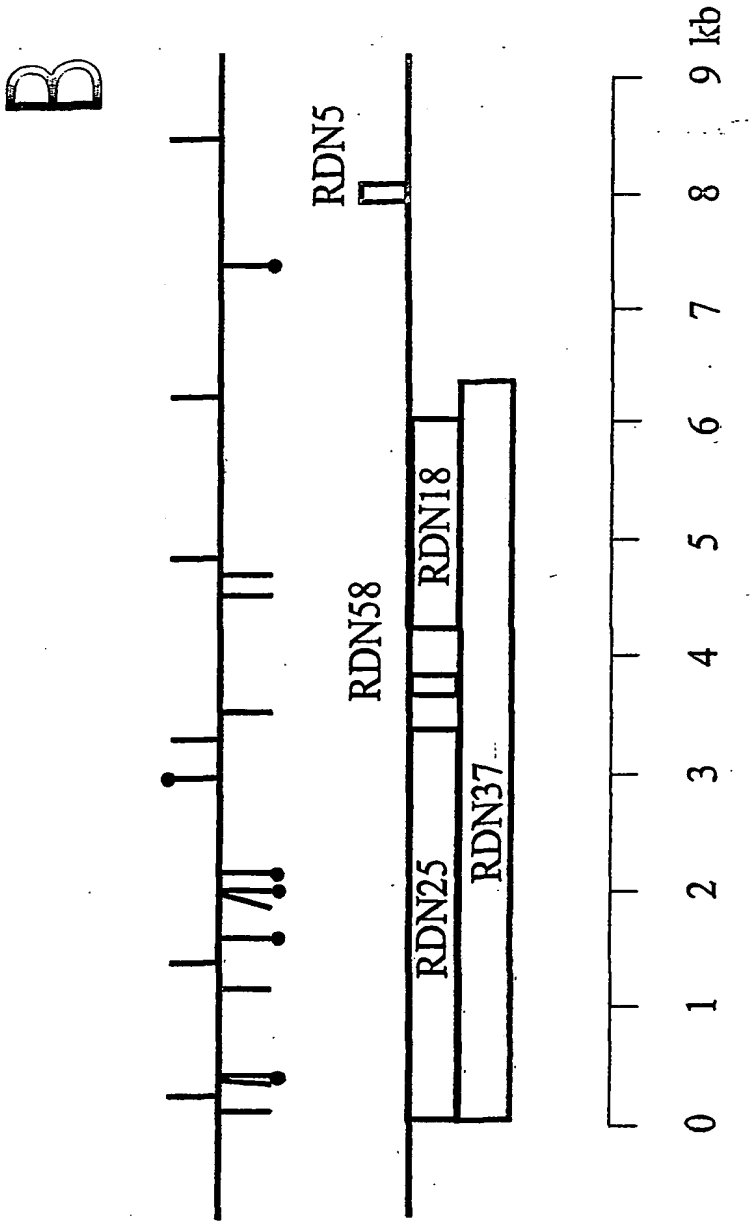


Fig. 5B

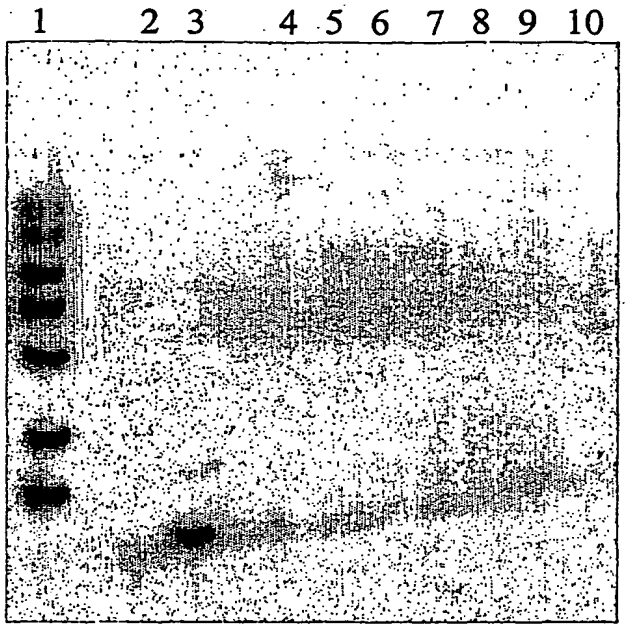


Fig. 6